PROVA GUIDES

CLAIM CLINIC

Activity | Outcome | Contribution | Prediction

Each kind of claim needs its own support

What Can We Responsibly Say?

A practical guide to explaining results with their important limits

FOR:
nonprofit leaders, program teams, boards, funders and colleagues preparing an account of results

USE WHEN:
a board paper, grant report, annual report or public statement needs a sentence people can stand behind

INCLUDES:
a completed workforce example, worked tables on who was counted and whose outcomes are missing, checks on missing outcomes, examples on participants' accounts and on uncertainty, practice with feedback and a claim record

HELPS WHEN YOU ARE:
explaining results, reviewing a public claim, telling an observation apart from a cause, or preparing for a board or funding conversation

This page is intentionally left blank.

Contents

01	The short answer	4
02	Identify the kind of claim	5
03	Recover the numbers behind the sentence	6
04	Who does the 70% describe?	7
05	Write the sentence the record supports	8
06	Can someone recover the support?	10
07	Explain how the program may have contributed	11
08	Notice what the missing answers could change	13
09	What could the missing answers change?	15
10	When a study supports a stronger statement	16
11	Practice on a different result	18
12	A 15-minute team exercise	20
13	A completed claim record	21
14	Use and review the explanation	22
15	Further reading	23

© 2026 Prova LLC · All rights reserved · PP-000073-10

01 The short answer

Your result from this guide: one revised statement, a short record of what supports it and where it stops, and a useful next check. You can complete a first pass even when the information you have is incomplete.

A report is nearly finished. Someone writes, “Our program helped 70% of participants find lasting employment.” The sentence sounds reasonable. Before publishing it, the team needs to know who was counted, what was asked, and whether the information supports “helped,” “find” and “lasting.”

Start with a sentence that names the people observed, the result measured and the period. Add the qualification that would change a reader’s understanding. Then check whether a claim about the program’s contribution needs more support.

If a statement is due today

1. Copy the exact sentence and name its audience.
2. Find the original source and the counts behind any percentage.
3. Name the people, the measure, the time and the comparison you actually observed.
4. Check each verb, and keep the important limit beside the result.
5. Ask a colleague what the revised sentence would lead them to believe.

Start with the workforce example and the completed claim record. Use the sections on contribution and missing outcomes for harder questions. The section on studies is optional. For a deadline, use the sheet What Can We Say Today?

© 2026 Prova LLC · All rights reserved · PP-000073-10

02 Identify the kind of claim

A board needs a clear account, and a funder needs accountability. Staff want their work recognized, and participants want their experiences told faithfully. These are reasonable needs. A careful explanation can meet them and still be precise about what is known.

Different sentences do different jobs. Counts of what was delivered, observed outcomes, explanations of change and predictions can each serve a worthwhile purpose.

Kind of claim	Example	What needs support
Activity	"We delivered 400 advice appointments."	The definition of the service, the count, the period and the quality of the records.
Observed outcome	"At six months, 56 of 80 respondents reported paid work."	The group observed, the question, the timing, the source and the missing responses.
Contribution or effect	"The program helped people obtain work."	A sound account of the program's role that considers other explanations.
Prediction	"Expansion will produce similar results."	The original finding, and whether the important conditions can be reproduced.

Keep the purpose in view.

If the audience needs an account of delivery, give that account clearly. If it needs to decide whether to expand, explain which part of that question the present information answers and which assumptions remain.

A claim about effect asks what might have happened without the program. This is called the **counterfactual**. A change from before to after does not, on its own, make that comparison. A well-designed comparison can help estimate an effect. Inquiry based on a theory of how the program works can look at how change happened and at competing explanations. A program does not have to be the only cause to make a difference. Further reading 1-3

Circle the words that carry weight: "reported," "improved," "helped," "caused," "proved" and "will." Ask what work each word is doing. "Improved" may describe a measured change. "The program improved" says the program was the cause.

© 2026 Prova LLC · All rights reserved · PP-000073-10

03 Recover the numbers behind the sentence

The people, places, organizations and figures in this document's examples are fictional.

Worked example. Northbank Skills is preparing a renewal report. Its proposed headline is: "Our program helped 70% of participants find lasting employment."

The source is a telephone follow-up six months after each person's first session, linked to intake and attendance records. The question asked whether people had done any paid work in the previous week. Follow-up was attempted with everyone who attended at least one session, including people who left before completing the service.

Record	Count
People enrolled	200
People who attended at least one session	160
People who completed the service	110
Starters who answered the six-month follow-up	80
Respondents who reported paid work	56

The table on who was counted, below, keeps every starter in view. Use it to check who the 70% describes.

The 70% is **56 divided by 80 respondents**. Half of the 160 starters answered. Among respondents, 60 completed and 20 left before completion. Completing the service and answering the follow-up are different facts. The result is not a percentage of completers, of all starters or of everyone enrolled.

The question measured paid work in one reference week. It did not show how the job was obtained, its pay or hours, or how long it lasted. This record has no comparison group and no measure of employment at the start.

© 2026 Prova LLC · All rights reserved · PP-000073-10

04 Who does the 70% describe?

Northbank route	Total people	Observed at six months	Not observed
Completed training	110	60	50
Started but did not complete	50	20	30
All starters	160	80	80

Among the 80 respondents, 56 reported paid work in the previous week and 24 did not. So the 70% describes respondents, and they cover 80/160 = 50% of starters. The 200 enrolled people, 160 starters, 110 completers and 80 respondents are different groups. There is no outcome for the forty people who enrolled but did not start.

© 2026 Prova LLC · All rights reserved · PP-000073-10

05 Write the sentence the record supports

The renewal report needs a clear result and enough context to interpret it. A sound version is:

> At six months, 56 of the 80 respondents reported doing paid work in the previous week. Respondents were half of the 160 people who started the service. Outcomes for the other 80 starters were unknown.

This gives the board a useful observation. Its qualification belongs in the main account, close enough to the result that it survives when the account is shortened.

Original wording	Reason for revision
“70% of participants”	The percentage describes respondents, not all starters or everyone enrolled.
“helped … find”	The follow-up records whether people were employed. It does not show the program’s part in that.
“lasting employment”	A question about one reference week does not show how long the job lasted or how good it was.

For a shorter chart label, use **“Paid work reported by 70% of six-month respondents”**, with the response count and coverage directly below it. A label such as “70% employment success” would leave the reader to infer too much.

The team notes two next questions: what happened to the people it did not reach, and what participants’ accounts and delivery records can tell it about the service’s contribution. It does not hold back

© 2026 Prova LLC · All rights reserved · PP-000073-10

Check the impression.
Ask someone who did not draft the report: “Who does this describe? What does it show about the service?” Revise if their answer is stronger than the information behind it.

urate descriptive report while it waits for answers.

© 2026 Prova LLC · All rights reserved · PP-000073-10

06 Can someone recover the support?

A citation to an institution or a polished chart does not let a colleague reproduce this result. Northbank's evidence file holds the extract of the intake group, the attendance extract, the follow-up question and the file recording each contact attempt and its result, and a calculation showing 56 / 80. The extract is dated 1 October 2026.

The person who owns the source checks that each starter has one unique ID, that the six-month timing rule was followed, and whether anyone answered twice. They also check that 56 yes, 24 no and 80 unknown add up to the 160 starters. The completed model assumes those checks were done within the case. Your worksheet must keep the checks that were actually done separate from the checks that are only proposed. Source 6

Check the route into the final product too. If a slide keeps only "70%," the repaired account has lost its denominator and its coverage again. In a chart, put **56 of 80 respondents, 80 of 160 starters observed** beside the result. In a short spoken account, say "70% of the respondents. We reached half of the starters." Keep the reference-week definition in the account that goes with it. If the format cannot carry the important limit, choose another format or a narrower claim.

© 2026 Prova LLC · All rights reserved · PP-000073-10

07 Explain how the program may have contributed

Suppose the workforce team also has three participants' accounts. Each person describes a coach's help with an application. Staff chose these people because they were willing to tell a positive story.

With permission and faithful context, those accounts can convey these people's experiences. They do not show how common that experience was. Because of how they were chosen, they also give little chance to find routes that did not work or competing explanations.

A stronger inquiry would ask how coaching was expected to help, and would look at evidence that could support or challenge that account. The table below is an original planning example. It is not a completed finding about cause.

Proposed explanation	Useful check	What could weaken it
Coaching improved applications.	Compare the accounts with dated drafts and application records.	The applications were already strong, or the job came before the coaching.
Participants used introductions to employers.	Trace the actual introductions, applications and hiring.	The jobs came through unrelated contacts.
Other conditions helped.	Ask about other services and about changes in local hiring.	The preferred explanation ignores these influences.

Ask whether the proposed evidence could look the same if the explanation were wrong. An application and a later job offer can appear whether coaching improved the application or the applicant already had the skills needed. A dated revision that shows a specific employer requirement being addressed tells you more about that particular link, especially when the applicant's account supports it and other influences have been checked. It still does not show an average effect on employment.

© 2026 Prova LLC · All rights reserved · PP-000073-10

A useful sentence for now:
"Three participants we chose described coaching that helped them prepare applications. These accounts point to a route worth investigating. They do not show how often it happens or the service's overall effect."

Include relevant experiences beyond the chosen success stories. Explain how people were chosen, how their accounts were examined and what contrary evidence was found. Any further interviews should have a clear purpose, and they should respect people's time and their right to decline.

Systematic qualitative inquiry can help investigate how a program works and what caused a change. Its strength depends on the design, the evidence and the reasoning. Changing "caused" to "contributed" does not supply that reasoning. Further reading 3-4

© 2026 Prova LLC · All rights reserved · PP-000073-10

08 Notice what the missing answers could change

Go back to the 160 starters. Eighty answered. Of those, 56 reported paid work and 24 did not. The other 80 outcomes are unknown. The employment rate among those 80 could be higher or lower than the rate among respondents.

Suppose the recorded answers are accurate, and every starter either did or did not meet the same definition of paid work at that time. Then two extreme cases show the size of the gap in information:

Assumption about the missing 80	Total in paid work	Share of all 160 starters
None met the definition.	56	35%
All met the definition.	136	85%

These are arithmetic limits under the stated assumptions. They are not predictions. They do not identify the most likely value, they do not allow for errors in reporting, and they do not estimate the program's effect.

A less extreme case can help a discussion. If 20% of the missing group met the definition, the rate for the whole group would be 45%: (56 + 16) / 160. If their rate matched the respondents' 70%, the rate for the whole group would also be 70%. The record alone cannot choose between those assumptions.

A response rate describes coverage. It does not tell us which way the result is biased or by how much. Using suitable information you already hold, compare the people who were reached with those who were missed. Look into what stopped people from being contacted,

© 2026 Prova LLC · All rights reserved · PP-000073-10

and ask whether more follow-up is practical and useful. Being similar on the recorded characteristics still leaves room for other differences. Source 2, chapter 9

Do not count an unanswered question as an unsuccessful outcome unless the measure clearly defines it that way and the account explains what that does to the result. Keep “unknown” visible when you report the result.

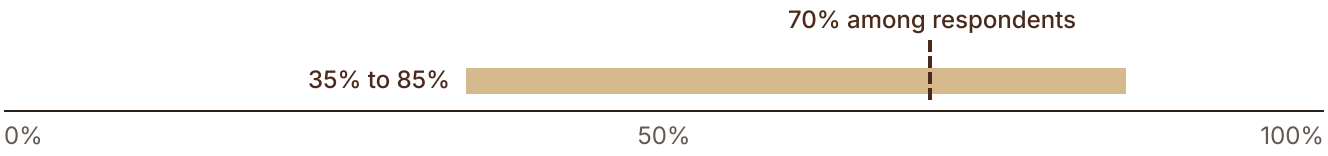
© 2026 Prova LLC · All rights reserved · PP-000073-10

09 What could the missing answers change?

Assumption about the 80 starters with unknown outcomes	Favorable outcomes among all 160 starters	Rate for all starters
None had paid work	56	35%
20% had paid work: 16 people	72	45%
70% had paid work: 56 people	112	70%
All had paid work	136	85%

The two extremes give arithmetic limits of 35% to 85%, as long as the known answers are correct and each one is a simple yes or no. The middle rows are assumptions. They are not estimates of the missing outcomes. None of these ranges is a range for the program's effect or a range of sampling error.

The whole-starter rate, depending on the 80 unknown outcomes



The range is arithmetic, not a forecast. The respondents' rate is one assumption about the missing group among many.

© 2026 Prova LLC · All rights reserved · PP-000073-10

10 When a study supports a stronger statement

Optional route: use this section when you are explaining a study's estimate of effect. You can complete the claim check without calculating an interval.

Example A. A summer-learning study offers a place to 120 children chosen at random and assigns 120 others to the usual provision. Follow-up scores are known for 114 children in each group. Assume that the way missing scores were handled justifies reading the result as the average effect of the offer for the assigned groups. Equal numbers observed in each group do not, on their own, justify that assumption. The estimated effect of the offer is 2 points on a 0-100 scale. Its approximate 95% confidence interval runs from -1.1 to +5.1.

The team had decided that 4 points would be worthwhile. That is a judgment for this case, not a standard for all of education. The interval includes both zero and 4. So the study cannot tell little benefit apart from a worthwhile improvement. Neither "no effect" nor "a worthwhile effect" is a supported conclusion.

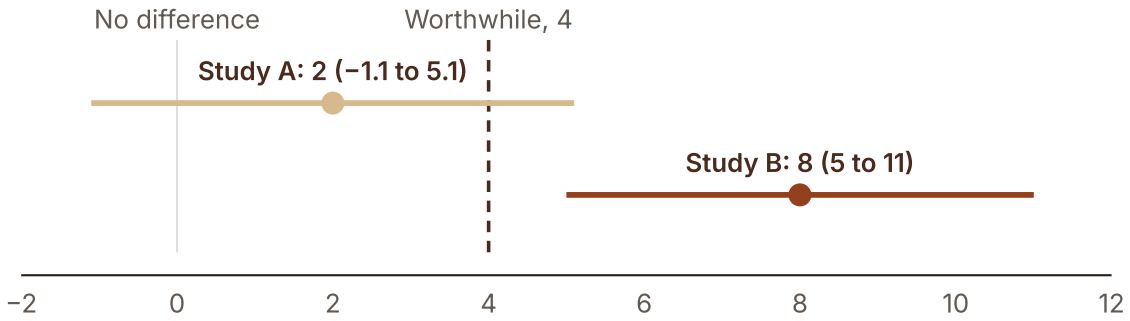
Example B: a favorable finding. Another study with random assignment has 120 children in each group and complete follow-up. Its estimated effect of the offer is 8 points, with an approximate 95% interval of 5 to 11. If the assignment, measurement and analysis are sound, it supports this statement:

> Offering a place increased the average follow-up score by an estimated 8 points compared with usual provision. The approximate 95% confidence interval was 5 to 11 points.

© 2026 Prova LLC · All rights reserved · PP-000073-10

The interval also lies above this team's 4-point benchmark. This supports a worthwhile average effect of the offer in this study. It does not show that every child benefited. It does not estimate the effect on children who attended, and it does not show what expansion would achieve. Further reading 1-2

Two estimated offer effects, in points



Study A's interval spans both no difference and the worthwhile benchmark. Study B's lies wholly above the benchmark.

A confidence interval describes precision under the assumptions of the design and analysis. It does not cover every possible bias. The 95% refers to how often the method for making the interval would capture the true effect across many comparable studies, if its assumptions hold. It is not a 95% probability that this one range contains the true effect. Source 5

These examples use a simplified analysis that compares two independent means. Example A also needs a sound way of handling the missing scores. Ask the author about the size of the effect, the uncertainty, the assumptions and the benchmark for what matters in practice.

© 2026 Prova LLC · All rights reserved · PP-000073-10

11 Practice on a different result

Housing example. One hundred households received support. At six months, the records show 52 households meeting the definition of housing stability and 22 not meeting it. The other 26 outcomes are unknown. For this exercise, the definition is housing that the household considers safe, with permission to stay for at least three more months. There is no comparison group. The target was that at least 80 of all 100 households would meet the definition.

Draft: “Our support made 70% of households stable, and missing responses mean we cannot tell whether we met our 80% target.”

Before reading the model, try three things:

1. Complete: “At six months, [count meeting the definition] of [count observed] households met [definition]. Outcomes for [count missing] were unknown.”
2. Explain what needs changing in “made 70% of households stable.”
3. Could the rate for the whole group reach 80% if every unknown outcome met the definition?

© 2026 Prova LLC · All rights reserved · PP-000073-10

Model and reasons

At six months, 52 of the 74 households with known outcomes met our definition of housing stability, about 70%. Outcomes for 26 households were unknown.

“Made” says the support caused the result, without any inquiry into cause to back it up. “Households with known outcomes” says whose results the percentage describes. The definition of the measure should go with the statement in the report.

If the recorded answers are accurate and each is a simple yes or no on the same definition, the rate for the whole group lies between 52% and 78%. Even if all 26 unknown outcomes met the definition, the total would be 78, below the target of 80. Here, the missing information leaves the exact rate unknown but still allows a conclusion about the target.

That conclusion concerns the defined target at this time. It does not show why the target was missed, whether the service helped or what should happen next. Different wording is fine if it keeps those distinctions.

© 2026 Prova LLC · All rights reserved · PP-000073-10

12 A 15-minute team exercise

Choose one real statement your team expects to use. Work from a suitable summary or a description of the source. You do not need files that identify participants for this exercise.

Minutes 0-2: Name the task. Write the exact sentence, the audience, the purpose and the deadline. Is the sentence reporting delivery, describing an outcome, explaining a contribution or recommending a next step?

Minutes 2-5: Recover the support. Identify the version of the source, the question or definition, the group, the period, the counts and the comparison. Mark any fact you cannot get as unknown. If the source cannot be checked now, record who will check it.

Minutes 5-9: Draft a revision. Write a sentence and the qualification that must go with it. Check every verb that carries weight. Keep the main result easy to read while keeping the limit that changes its meaning.

Minutes 9-12: Challenge the impression. Ask a colleague to explain who is represented and what the statement shows. Invite one plausible other explanation or problem with the measure. Use the guide's examples when the group needs a model.

Minutes 12-15: Record what follows. Complete the claim record. Name one useful check, who owns it, and the person or body responsible for approving the account. If more work is needed, record a sound statement to use in the meantime, or hold back the claim that is not supported.

The exercise can produce a first pass, not a final decision for the organization. If colleagues disagree, work out whether they differ about the facts, the interpretation, what the audience needs, priorities or who has authority.

Success looks like:

a sentence that another reader interprets correctly, a source that can be checked, unknowns kept in view, and a next step with an owner.

© 2026 Prova LLC · All rights reserved · PP-000073-10

13 A completed claim record

This record shows what the workforce team takes into the review of its report. Use a blank copy of the claim record for your own statement.

Field	Completed example
Audience and purpose	The foundation's renewal panel. The purpose is to describe employment at six months.
Original statement	"Our program helped 70% of participants find lasting employment."
Source and definition	Northbank's extract of 1 October 2026, linking intake, attendance and the six-month telephone follow-up by starter ID. The measure is any paid work in the previous week. The checks that the numbers add up and that timing was followed were completed for this model.
Group and counts	200 enrolled, 160 started and 110 completed. 80 answered (60 completers and 20 other starters), and 56 reported paid work.
Comparison and contribution	There is no measure of employment at the start and no comparison group. Three chosen accounts suggest a route worth investigating.
Revised statement	At six months, 56 of 80 respondents reported paid work in the previous week. Respondents were half of the 160 starters.
Essential qualification	Outcomes for the other 80 starters were unknown. How long jobs lasted and how good they were was not measured.
Other explanation or gap	The people reached may differ from those missed. Other services and local hiring conditions may matter.
Next check and owner	The evaluation lead checks coverage and whether more follow-up is practical, before the next review.
Review and authority	The program manager checks the definitions and counts. The executive director approves the account for the renewal.

Checking the source, interpreting the result and approving the account for the organization are separate responsibilities. Follow your organization's own arrangements for approval.

© 2026 Prova LLC · All rights reserved · PP-000073-10

14 Use and review the explanation

Keep the result and its important qualification together in the board paper, slide, chart caption or grant report. A note on sources can carry more detail, but a central limit should survive when people quote or shorten the account.

Come back to it later:
bring a fresh sentence to the next team meeting. Before reopening this guide, identify its group, its measure, its period and any language about cause or prediction. Compare your revision with a colleague's and explain any difference.

Before use, confirm the counts and definitions with the person responsible for the source. Check that quoted experiences are told faithfully and that people have agreed to how they will be used. Record the version, the review date, any disagreement still open and who owns the approval. If a denominator or definition changes later, explain the change before comparing results over time.

Ask for specialist review when the statement depends on a technical analysis of cause, statistical uncertainty the team is not familiar with, disputed definitions, or consequences that the review you have cannot properly assess. Say exactly which question needs review.

© 2026 Prova LLC · All rights reserved · PP-000073-10

15 Further reading

1 **U.S. Government Accountability Office (2012). Designing Evaluations, GAO-12-208G.** Chapter 3, especially printed pages 18-20, connects questions, information, methods, limitations and conclusions. Publication and report.

2 **Gertler, Martinez, Premand, Rawlings and Vermeersch (2016). Impact Evaluation in Practice, second edition. World Bank and IDB.** Chapters 3-4 cover comparisons with what would have happened otherwise, and how people are assigned. Chapter 9 covers people lost to follow-up and other challenges of method. Chapter 15 covers precision and sampling. Publisher overview.

3 **HM Treasury and Evaluation Task Force (2026). Magenta Book, Annex A.** Sections A1.3-A1.4 discuss process tracing and contribution analysis, including competing explanations and limitations. Full guidance.

4 **Spencer, Ritchie, Lewis and Dillon (2003), republished by HM Treasury (2012). Quality in Qualitative Evaluation.** The framework's items on design, sampling, analysis and credible findings. 2012 edition. Official web version, updated in 2026.

5 **Greenland and colleagues (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European Journal of Epidemiology, 31, 337-350.** See the discussion of confidence intervals, and of statistical conclusions compared with conclusions about what matters in practice. Full article.

6 **U.S. Government Accountability Office (2019). Assessing Data Reliability, GAO-20-283G.** How to judge whether data suit the intended use, trace data back to source documents, and record the checks and limitations. Publication and report.

The frameworks in this guide are Prova's own syntheses.

Notes on method

The limits in the workforce and housing examples use one fixed yes-or-no definition and assume that the known answers are correct. They describe the unknown outcomes in the whole groups stated. They are not ranges for the program's effect or ranges of sampling error. The learning examples use a standard deviation of 12 in each group. A: standard error = sqrt(144/114 + 144/114), about 1.59, and interval = 2 +/- 1.96 × 1.59. B: standard error = sqrt(144/120 + 144/120), about 1.55, and interval = 8 +/- 1.96 × 1.55, rounded to 5-11. The effect of offering a place is different from the effect of attending.

© 2026 Prova LLC · All rights reserved · PP-000073-10

QUESTION IT

Found an error, or want to question this document? Contact Prova at hello@byprova.com.

© 2026 Prova LLC · All rights reserved

Cite as: Prova (2026). What Can We Responsibly Say? byprova.com

Examples are fictional unless sourced. Not legal or financial advice.