Can I make this policy work here?

Better theories of change and better evidence management for better policy design, better policy prediction

Illustrated with a South African HIV/AIDs policy intervention

Nancy Cartwright, Durham University
Brendan Kelters, Durham University
Eileen Munro, London School of Economics
Likhwa Ncube, University of Johannesburg
John Pemberton, Durham University

CHESS Working Paper No. 2026-01 Durham University

CENTRE FOR HUMANITIES
CHESS
ENGAGING SCIENCE AND SOCIETY

Chess working paper (Online) ISSN 2053-2660

Can I make this policy work here?

Better theories of change and better evidence management for better policy design, better policy prediction

Illustrated with a South African HIV/AIDs policy intervention

Nancy Cartwright, Durham University
Brendan Kelters, Durham University
Eileen Munro, London School of Economics
Likhwa Ncube, University of Johannesburg
John Pemberton, Durham University

Professor Nancy Cartwright
Centre for Humanities Engaging Science and Society (CHESS),
Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK
nancy.cartwright@durham.ac.uk

Dr Brendan Kelters
Centre for Humanities Engaging Science and Society (CHESS),
Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK
brendan.t.kelters@durham.ac.uk

Professor Eileen Munro
Department of Social Policy, London School of Economics, 2nd Floor, Old Building, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE e.munro@lse.ac.uk

Dr Likhwa Ncube
Centre for Philosophy of Epidemiology, Medicine, and Public Health (CPEMPH),
Faculty of Humanities, University of Johannesburg, PO Box 524, Auckland Park, 2006, South Africa
likhwan@uj.ac.za

Dr John Pemberton
Centre for Humanities Engaging Science and Society (CHESS),
Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN, UK
john.m.pemberton@durham.ac.uk

This guide is a version of Durham University's CHESS Working Paper No.2025-1, with a different running example. Both guides are overviews of the methods proposed in Cartwright, N., Munro, E. and Pemberton, J. (forthcoming), *Causal Processes and their Warrant: A Practical Guide*. Cambridge: Cambridge University Press. This work was supported by the Arts and Humanities Research Council (Grant Ref: AH/X006727/1). It was also supported by British Academy South Africa grant *Integrating Evidence for Contextualised Public Health Policy: Lessons from South Africa* (Grant Number: EPG\100488).

Quick Start Guide to the Evidence Role Map Framework (ERMF)

Will this policy work here?

What is ERMF?

While social systems are dauntingly complex to study, this framework offers a pragmatic way to tackle the challenges of deciding whether a policy can help you achieve your goals. It is a structured way to think about how your policy will work or whether it has worked, what evidence is needed and how to organise your evidence. It aims to help you:

- Improve policy design and implementation
- Make more reliable predictions about success *in this setting*
- Support both post-hoc evaluation and mid-course correction during delivery.

Who is ERMF for?

Implementation planners, programme teams, monitoring and evaluation staff, technical policy analysts and advisers. (For a more detailed account of the approach, see Cartwright, Munro and Pemberton (2026), *Causal Processes and their Warrant*, Cambridge University Press.)

Deliverables

Using the ERMF will give you:

1. A **Theory of Change (ToC)** diagramming a rich account of how the policy is expected to achieve its goals and what can help or hinder this in your context
2. A catalogue of **evidence** to estimate whether each step in the ToC can be achieved
3. One or more **evidence-role maps** (short "diagrammed arguments" showing how evidence supports or undermines key claims)
4. Help in identifying **key gaps** in your evidence and in your proposed design fixes
5. A **qualitative confidence judgement** (strong / medium / weak) about how feasible it is to implement the ToC or how well it has been done in your context.

Seven Key Facts for any ToC step (the checklist)

For a step to work *as envisaged*, assess whether:

1. **The proposed cause will occur** (at the right time/place/scale).
2. **The proposed effect will occur** (at the right time/place/scale).
3. **No other sufficient cause makes the step redundant** (the effect is not mainly produced anyway).
4. **Required support factors will be in place** (conditions needed for success).
5. **Derailers/detractors are absent or safeguarded against** (or their impact is managed).
6. **The activity** by which the cause is to produce the effect **will carry through start-to-finish, under workable principles** (the mechanism is feasible here).
7. **The underlying system can support the activity/principles** (stable institutional/social/logistical conditions).

Rules of thumb

- **No evidence is not evidence against.** Mark “unknown” explicitly.
- **Treat the ToC as revisable:** the goal is not to defend it, but to make it more accurate and usable.
- Prefer *explicit reasoning* over false precision: “strong/medium/weak” is often the right level.
- Prioritise evidence collection on the **weakest links** and the **most modifiable conditions**.
- If right at the start it seems unlikely that you will be able to fix what worries you**, consider adopting a different plan.**

If briefing senior decision-makers

Use the ToC diagram accompanied by 3 bullets: (i) key risks, (ii) key mitigations, (iii) biggest evidence gaps.

What we do here

We tackle a well-known pitfall in RCT-based policy evidence: in real life what works in studied populations often fails to work elsewhere. Context makes all the difference. What we offer is a way of studying your context and considering whether it will, or can be modified to, support the policy. Of course, if there have been several studies of the policy in a range of settings, their findings will be helpful, especially if those settings are *similar to* yours in the right ways. But how do you know what makes for 'the right ways?' What is a *relevant* similarity? Often, the reason for doing an RCT in the first place is ignorance about the range of factors that can affect whether the policy will produce the targeted outcomes. Also, much of the time, you are using or planning a policy which has not been subjected to an RCT so how can you judge both its feasibility and likely effectiveness in your setting?

We contend that the best starting point is to look to your setting itself to try to understand how the policy would work there. You need to ask: 'How would the policy lead to the intended outcome here?' i.e. 'What is the causal pathway?' 'What's needed for this pathway to operate start-to-finish here in our setting?' or for post hoc evaluation, 'How has it operated?'

But what should you focus on? What *kinds* of facts about the process, setting and the way the policy will be implemented matter for describing causal pathways and establishing whether they will carry through? We offer an account of these facts in Part 1. Part 2 is to help you organise the evidence you collect about these facts in a way that makes clear the role that each piece of evidence plays and where more evidence is needed. Together this rich step-by-step account of what is needed for a policy to work and of the evidence about whether these requirements are met should put you in a stronger position from which to plan, tailor and implement policy.

In summary, this is a guide on using evidence for

- Better policy design and implementation
- More reliable prediction about whether a policy will work in your setting
- More accurate post hoc evaluation of whether a policy has worked there as intended and a deeper understanding of how it has worked.

This guide will show you how to

- Develop an *information-rich theory of change* showing the process by which your policy design is supposed to achieve its targeted outcomes – *how* the proposed interventions are to lead step-by-step to the hoped-for outcomes

- Classify and use *evidence* about whether your policy is likely to work if implemented as planned
- Develop a rough overall assessment of how well justified you are in believing that the process you envisage will go through start to finish
- *Leverage* such assessments to predict how likely you are to be able to implement a policy successfully
- *Assess* how well a policy is working to allow you to fix things that are not proceeding properly, or abort the policy before too much is invested if it is failing
- Use such assessments to *enhance policy design and implementation*.

A Running Example

We use one example throughout, a South African HIV-AIDS programme, chosen both because it is a real case for which a post hoc evaluation is available and because it is easy to understand and not too complex to diagram. Since our primary focus is on policy planning, we shall discuss this policy from a hypothetical *ex ante* perspective. We will illustrate how a rich theory of change (ToC) could be developed and evidenced for this policy, discussing how this ToC can be used to assess and improve chances of policy success.

The policy we focus on involved providing PrEP (short for 'pre-exposure prophylaxis') free to female sex workers (*FSWs*) in eThekwini, South Africa, 2016-2020, in the hopes of securing a high level of uptake and persistent use and thus reducing their risk of developing HIV. The programme started in clinics serving sex workers. It later expanded to other groups, with educational materials to promote awareness and help individuals assess if PrEP would be right for them (Pillay et al 2020; doi:10.1371/journal.pone.0228620). The medication was dispensed, after clinical assessments by staff working at clinics run by the organisation TB HIV Care.

Based on a post hoc evaluation by Amrita Rao et al, published in March 2022 (doi:10.1371/journal.pone.0265434), the policy was only modestly successful. Here, we present an information-rich ToC to help understand why it did not go as well as hoped which, done before implementation, might have helped predict this underperformance and we suggest improvements in the policy's design.[1]

Our overall purpose

Many of the considerations we raise will seem obvious, but they become much more useful if you make them *explicit* and *assemble them together* in a procedure that makes each one harder to miss. After all, policy failure is common and, after the fact, it is often attributed to things that could have been realised beforehand.

[1] Some such improvements were indeed implemented in this example; namely the addition of extra resources to contact FSWs and remind them about follow-up appointments.

The framework helps you think about and keep track all at once of the large variety of things that might matter to success, including things you might need to change to secure success. In support of this goal, we'll show you how to organise your evidence about these things in a way that helps you see what each piece of evidence is doing to better estimate how your total body of evidence supports – or does not – a prediction of likely policy success.

Part 1 describes how to construct an information-rich ToC that depicts how you think your policy interventions will lead to your hoped-for outcomes and what you think is needed for this to happen. Part 2 describes how to use your information-rich ToC to catalogue the evidence you collect and understand what roles it plays in supporting – or undermining – a claim of policy success. We conclude in Part 3 with a brief summary.

Part 1: Constructing a rich theory of change

A step-by-step account of how a policy works is called the policy's **theory of change (ToC)**. It details the **successive steps** by which the policy is supposed to lead to the targeted outcomes.

The standard advice is to develop a **boxes-and-arrows diagram** that shows the main sequence of steps (these are also sometimes called 'logic models'). We start there too but then explain how to enrich this to give more detail about *how* you get from the very first policy interventions to the final targeted outcomes: what helps or hinders each step, and how these steps depend on the local context. This enriched ToC will be a practical tool for **design, implementation and evaluation** of the policy.

Constructing a basic boxes-and-arrows ToC

Basic boxes-and-arrows theories of change are very familiar now and are widely employed in policy planning and evaluation**.** See Figure 1 for an example we have constructed, depicting a way in which the provision of monthly PrEP free to FSWs at TB HIV Care clinics in eThekwini could be expected to lead to a high level of uptake and persistent use among them, where persistence is defined as year-long use, which is what you see recorded in Box 13 as the hoped-for outcome.

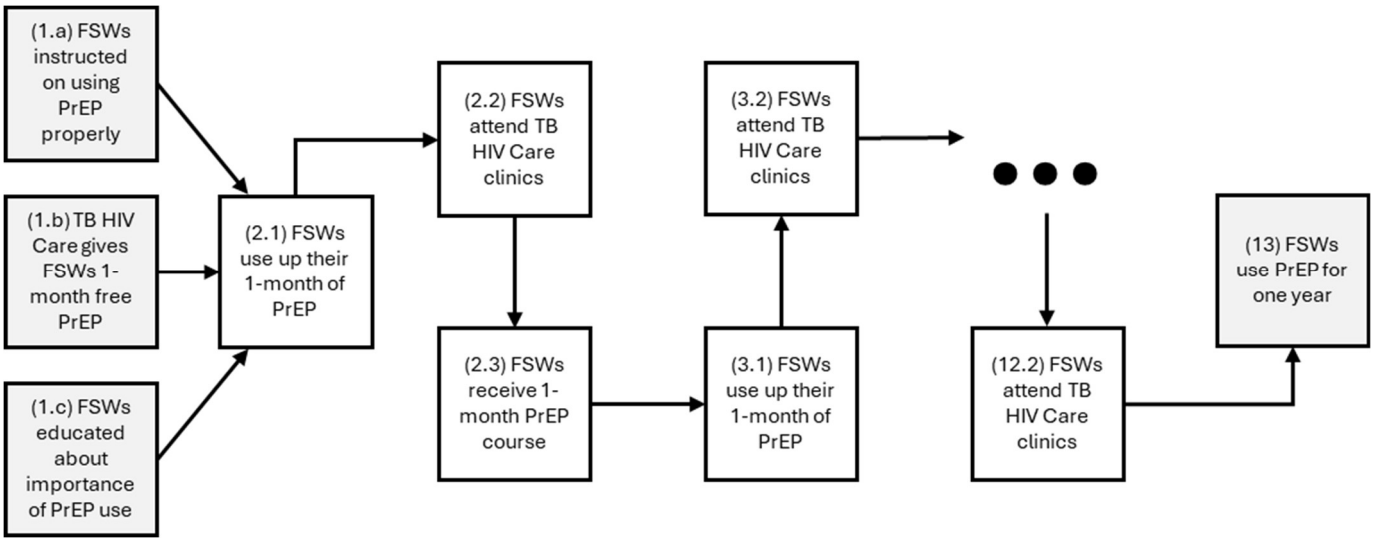


Figure 1: Example of a basic Boxes-and-Arrows Diagram.

Figure 1 is an example of a ToC in the usual form, with the proposed **policy interventions** at the start, the **desired outcome** at the end, and in between a series of **intermediate** cause-and-effect steps by which the interventions are expected to contribute to achieving the outcomes**.**

In this case, the policy involves three separate interventions at the first step, each of which is expected to encourage PrEP use. FSWs are not only to be given a 1-month dose of PrEP free (1.b), but they are also to be instructed in its use (1.a) and educated about the importance of using it (1.c). The desired outcome is year-long use of PrEP among the FSWs who start the programme. Notice that after the first inauguration, the same three steps are expected to happen each month to, in the end, ensure year-long use. Since, from the 2nd month through the 12th these steps are the same, we will not talk about each separately but all of them together using the label *N* (N = 1,…,13)

We will now show how to enrich such boxes-and-arrows ToCs by filling in *what is supposed to happen at each step to ensure that it will lead to the next*. This richer ToC can then be the basis of further exploration that can help you develop and analyse evidence that is useful for figuring out a policy's chance of success in your setting and what can be done to maximise that. Recognising problems at any stage, you may be able to suggest policy design changes to ameliorate them. If this can't be done you should probably reduce your estimate of the chance that the policy will have its intended effect.

Enriching a boxes-and-arrows ToC

The boxes in a standard boxes-and-arrows ToC generally have short labels, so the first step is to describe them more fully – the more you understand about how the features in each box are expected to affect the process, the better you will be able to see if it can work, and how. Include here expected **timings** for a step to lead to the effect, the **magnitudes** of the impact and their **locations** (where, or to what or whom, do you expect the effects and causes to happen). This helps you think through whether what needs to happen at each step matches your best estimate of what is likely to happen there.

Next, consider for each step just **how** that step is supposed to occur. **What must happen** for the cause at that step to produce the expected effect, and **what might help or hinder this?**

We suggest you think in terms of these categories:

- **Support factors** that are needed if the cause at the step is to produce the effect.
- **Derailers** that *stop* the cause from producing the effect and **detractors** that *hinder* the cause producing the effect, reducing its contribution.
- **Safeguards** that lessen the effects of derailers and detractors.
- The **activities** by which the cause produces its effects.
- The **principles** under which these activities occur in settings like this.
- Characteristics of the social/economic/legal/cultural/geographical/physical/etc. **underlying systems** that allow these principles to obtain and the related activities to happen.

Support factors

Causes rarely work alone. For example, consider Box N.1 ('FSWs use up their 1-month of PrEP') in Figure 1. We cannot expect this to generally lead to Box N.2 ('FSWs attend TB HIV Care clinics'), unless at least two further facts hold:

> S1(N.1-N.2) FSWs find the clinic accessible enough
> S2(N.1-N.2) FSWs don't fear returning too much (e.g. due to social stigma of being seen at the clinic or violent reactions from their partners).

If either of these fails for a substantial number of FSWs, then you cannot expect high levels of FSWs returning to the clinics to receive follow-up doses. (Note that we label support factors with *S* followed by the name of the step they are meant to be a support factor for.)

Derailers and detractors

These are respectively conditions that if present prevent a cause and its support factors from bringing about its effect or reduce the contribution they make. For example, here are two possible detractors for S1(N.1-N.2):

> D1/S1(N.1-N.2) FSWs must travel long distances to the clinic
>
> D2/S1(N.1-N.2) FSWs expect long waiting times at the clinic.

And here's a possible detractor for S2(N.1-N.2):

> D1/S2(N.1-N.2) FSWs fear they will be publicly visible waiting in queues outside the clinic.

(Detractors/derailers are labelled *D* separated by a slash from the feature they detract from.)

Safeguards

Safeguards are conditions which thwart derailers and detractors. For instance, here is a plausible guard against detractor D1/S2(N.1-N.2):

> G1/D1/S2(N.1-N.2) Ample space is provided in a private waiting room inside the clinic.

Of course, even when you have envisaged a derailer or detractor, it's sometimes impractical to guard against it, in which case you will need to assess your chances of policy success as lower. How much lower depends on your assessment of the likelihood of the detractor or derailer occurring as well as on how strongly the detractor detracts.

(Guards are labelled by *G* separated by a slash from the factor they guard against.)

There is obvious interplay between support factors, detractors/derailers and safeguards. For example, the absence of a support factor may be thought of as a derailer, or a safeguard against a derailer may be thought of as a support factor in a ToC that omits this derailer. We recommend not getting caught up in deciding which categories to place your factors in. Just concentrate on making sure you've accounted for each somewhere, without duplication.

Activities

These are the actions by which causes produce their effects. It is important to think through what the activity is supposed to be at each step since this provides clues as to what support factors will be needed and what might detract from or safeguard the process. This might yet again seem trivial but attending to *how* one stage is to lead to another can alert you to features you might otherwise miss. For instance, in the free-PrEP intervention in eThekwini, an FSW's using up all their PrEP by the end of the month was presumably supposed to do this:

A1(N.1-N.2) *motivate* FSWs to return to restock.

(Activities are labelled by *A* followed by the step at which they occur.)

This suggests the importance of support factors S1(N.1-N.2) and S2(N.1-N.2) since difficulties in accessing the clinic or fear of returning can be seriously demotivating. It also suggests another possible support factor:

S3(N.1-N.2) FSWs who miss appointments are called to encourage them to attend.

We know that this support factor was reasonable to consider in designing the free-PrEP programme since it was in fact put in place in 2019, with the return rate improving thereafter. S3(N.1-N.2) itself was the effect of a new policy initiative in June 2019, which, in our terms, required at least:

S1/S3(N.1-N.2) Addition of new staff, including case managers and social workers.

This is all supposing the relevant activity from N.1 to N.2 is 'motivating'. Other activities might have been at play instead if the circumstances were different. For instance, in a very prescriptive legal system, running out of PrEP could *legally compel* FSWs to return to the clinic, in which case support factors would include FSWs recognising and understanding that law and being motivated to obey it. Or it could be that running out of PrEP *initiates a process of physically conveying* FSWs to the clinic. This would make 'The means of conveyance is seen by FSWs as sufficiently efficient' a support factor, along with 'The means of conveyance are available', for N.1 to cause N.2.

Another activity by which N.1 can be expected to lead to N.2, though probably less strongly than A1(N.1-N.2), is A2(N.1-N.2) is *reminding*. That is, the fact of finishing the PrEP dose serves as a reminder to return for more, an activity that can also be stimulated by S3(N.1-N.2).

One giveaway for recognising activities is that they tend to be named by verbs.

Principles

We assume that it is not arbitrary which activities a cause can initiate and which outcomes can be produced from these; at least, it is not arbitrary in cases where you can hope to predict or explain. There is some 'systematicity' to it: the activities happen in accord with principles that we can learn and learn how to use. These principles describe the general patterns that the activity exemplifies. Many are familiar everyday principles that we regularly appeal to in explaining and predicting what happens, like 'People avoid actions they expect to get punished for', 'Parents act to protect the welfare of their children' and 'People act to maximise their expected utility'.

These are often called ‘tendency’ principles because they need not describe what follows when a cause acts but rather what the cause ‘tends’ to do. Often, what a cause ‘tends’ to do is not the same as what actually happens because other causes operate as well, and what actually happens depends on all of them together. (What we call ‘tendency principles’ are sometimes also called ‘mechanisms’, especially in Realist Evaluation.)

Even when you know what activity is supposed to occur, knowing the principle under which it occurs can be of further use in identifying support factors and derailers. For instance, the activity ‘motivates FSWs to return to restock’ could be expected to take place under this principle:

> P1/A1(N.1-N.2) People are motivated to avoid harms to themselves.

(Principle are labelled by *P* followed by the activity to which they apply.)

Or alternatively under this principle:

> P2/A1(N.1-N.2) People are motivated to engage with services when they are free, non-judgmental and personalised.

Or, as we imagine is expected to be the case, both. If the first, as a support factor for Box N.1 to lead to Box N.2, the original messaging (represented in Box 1.c) should stress, e.g., the horrors of contracting AIDS, the higher risk of FSWs contracting HIV and the remarkable protection appropriate PrEP usage can provide. If the second, the messaging should emphasise that the PrEP doses are free, provided without judgment and that each client will be treated with respect as an individual with personalised care and service.

Variations in the principle under which the motivation activity worked might have mattered a great deal in the eThekwini programme. Rao et al (2022; doi:10.1371/journal.pone.0265434, p. 4) note that “Younger women were more likely to discontinue PrEP due to missed visits compared with those 25 years and older”, attributing this in part to “lower risk perception” in the young. This suggests less motivation, specifically under principle P1/A1(N.1-N.2) amongst this group.[2]

Underlying systems

‘Underlying system’ is the label we give to the relatively stable social, economic, legal, cultural, geographical, physical and so on arrangements that obtain in the setting. These are the kinds of arrangements that are the bread-and-butter of much of contemporary systems theory. Here, we are concerned with all those relatively stable background conditions that a realistic implementation of your policy has little chance of altering. What kinds of causal relations are afforded by these underlying arrangements, and what kinds are forestalled?

The settings in which free PrEP programmes for FSWs are introduced typically provide a legal framework, clinic and outreach infrastructure, facts of HIV orientation, resource availability and a particular economic and political backdrop, as well as various cultural and social norms

[2] The study suggests this reason along with others like “lower levels of HIV-related knowledge…less experience in the [sex] industry, less social support and fewer network ties, greater vulnerability to violence, lack of financial autonomy, and underdeveloped self-efficacy” (2022, p. 6).

and expectations, all of which can help afford or inhibit the activities and principles that policy success depends upon. For example, one underlying condition that seems to have obtained is

> U1(N.2-N.3): Reliable stock and quick dispensing, especially for HIV and AIDS prevention and protection, is a routine service provision in many public health facilities in South Africa.

(Features of the underlying system are labelled by *U* followed by a label for what they are taken to help afford/inhibit.)

This is presumably important in ensuring that PrEP treatments are available at the clinic for those who return, which is a major support factor for the step from N.2 ('FSWs attend TB HIV Care clinics') to lead to Box N.3 (FSWs receive 1-month PrEP course'). Another condition important to this, reported by Rao et al, seems to be

> U2(N.2-N.3): HIV prevention and treatment services remained "essential" in South Africa during the COVID pandemic (Rao et al 2022; doi:10.1371/journal.pone.0265434, p. 6).

Of course, the COVID pandemic could not have been predicted, but clearly, foreseeing whether enough doses will be at the clinics at the required times is important to foreseeing whether high levels of persistence will be achieved.

Another relatively stable system feature described in the Rao et al study that matters is one we mentioned above, which we here label U1(N.1-N.2).

> U1(N.1-N.2): Young FSWs tend to have "underdeveloped self-efficacy" (Rao et al 2022; doi:10.1371/journal.pone.0265434 p. 6).

This feature clearly reduces the applicability among younger FSWs of P1/A1(N.1-N.2): People are motivated to avoid harms to themselves.

Another underlying condition is

> U2(N.1-N.2): There is a fair amount of social stigma associated with taking PrEP in South Africa(Pillay et al 2020; doi:10.1371/journal.pone.0228620)

which inhibits the success of the motivation activity, whether motivation is to operate under P1/A1(N.1-N.2) or under P2/A1(N.1-N.2).

Generally, if the underlying system supports some principle by which a step's cause tends to bring about its effect, then the step *can* work (though it may not if all else is not in place properly). If the system doesn't support any such principle, the step will fail in that setting under any realistic conditions.

There is no clear-cut line between what is 'underlying' and what is part of the causal process you are focusing on. A helpful strategy for deciding whether you should count a factor as part of the underlying system or as part of your intended process is to consider *ease of changeability.* If the current system is missing factors necessary for a step to go through, can they reasonably be changed, or would that require many factors in the setting all to change together in an integrated way? Since systemic change is so much harder, the sooner you recognise that it

might be needed, the sooner you can get started on it or decide it is too hard and begin considering alternatives.

Diagramming an enriched ToC

Figure 2 illustrates how step N.1 can lead to step N.2 or any step N after the first.

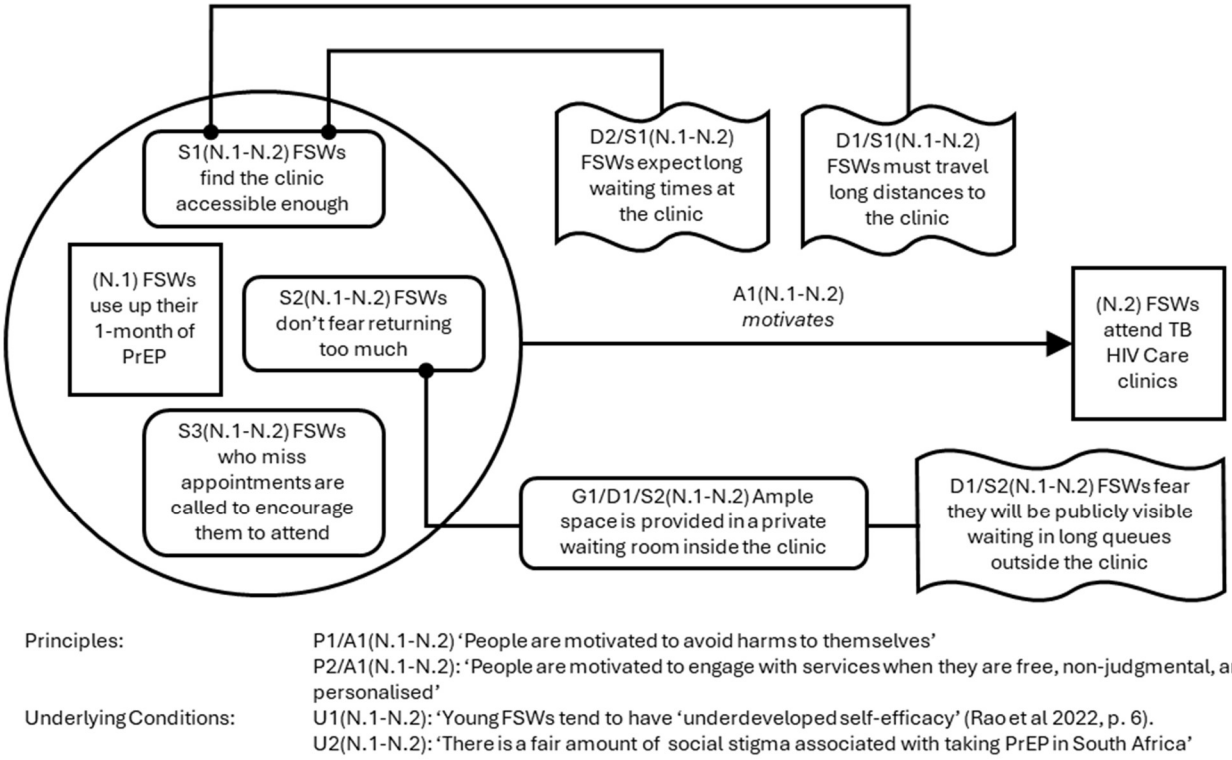


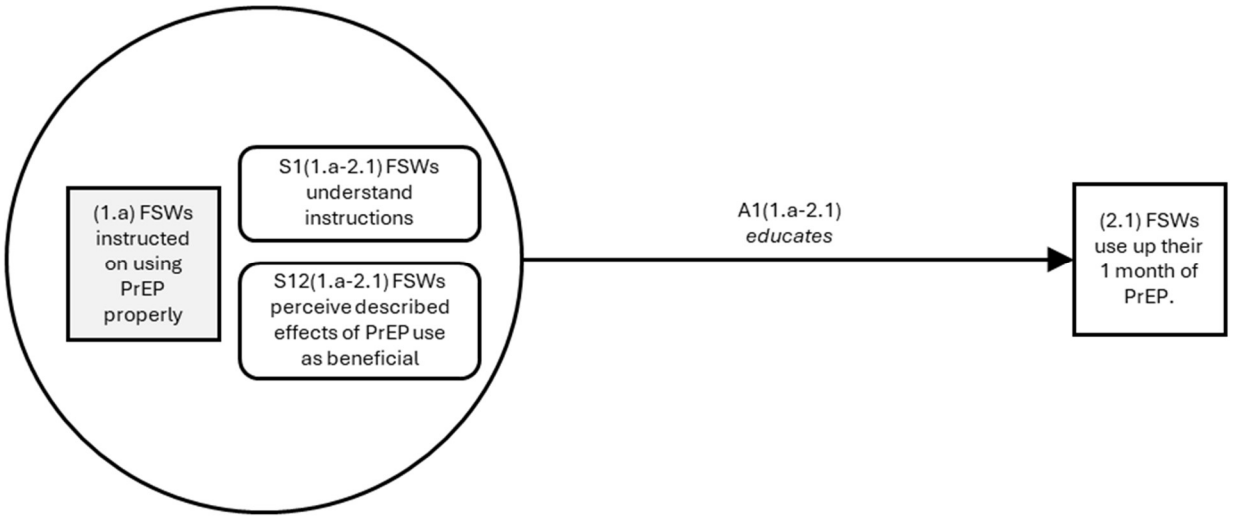
Figure 2: A Causal Step from a Diagram for an Enriched Theory of Change, detailing N.1 →N.2 from Figure 1.

Notice that the factors that are needed to act together to produce a result, including what was thought of as the main cause, are enclosed in a circle with the tail of the arrow going from the edge of the circle (since it is all these factors together that are needed to produce the result). Those that act independently of each other have separate arrows, as in Figure 1. All the factors within a circle are meant to be equally necessary for producing the effect. Nevertheless generally one is the most salient, like the ones depicted boxes in simple boxes-and-arrows diagrams. So we distinguish what are normally thought of as support factors from those by placing support factors in boxes with rounded corners. Detractors are graphed in boxes with squiggly edges and safeguards in boxes with rounded edges placed between the detractors and the factor that is guarded.

The step represented in Figure 2 appears as only one of the recurring steps in the larger diagram of Figure 1.

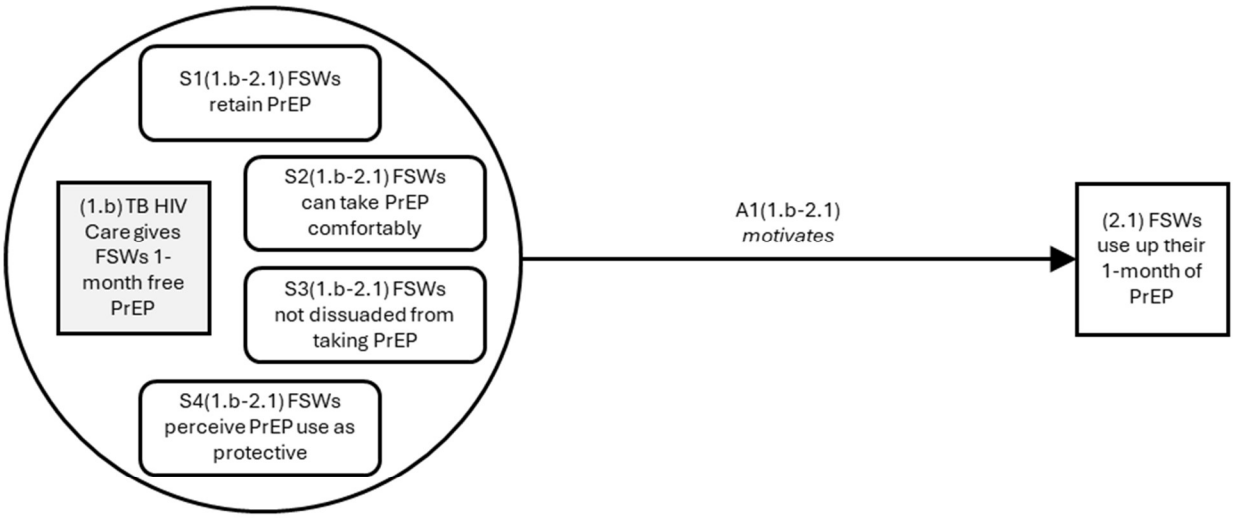
For further illustration, we'll now add some of the same sort of additional details for the moves from Boxes 1.a, 1.b and 1.c to Box 2.1, and from Boxes N.2 to N.3 and N.3 to [N+1].1. In these diagrams, as in Figure 2, many real and relevant details are omitted. This is inevitable in actual

practice, due to the limited resources available for planning and the complexity of what you're investigating, but you should aim to try and capture and describe at least the more impactful factors and activities. The move from 12.2 to the outcome represented in Box 13 (see Figure 1) is just a matter of accounting.



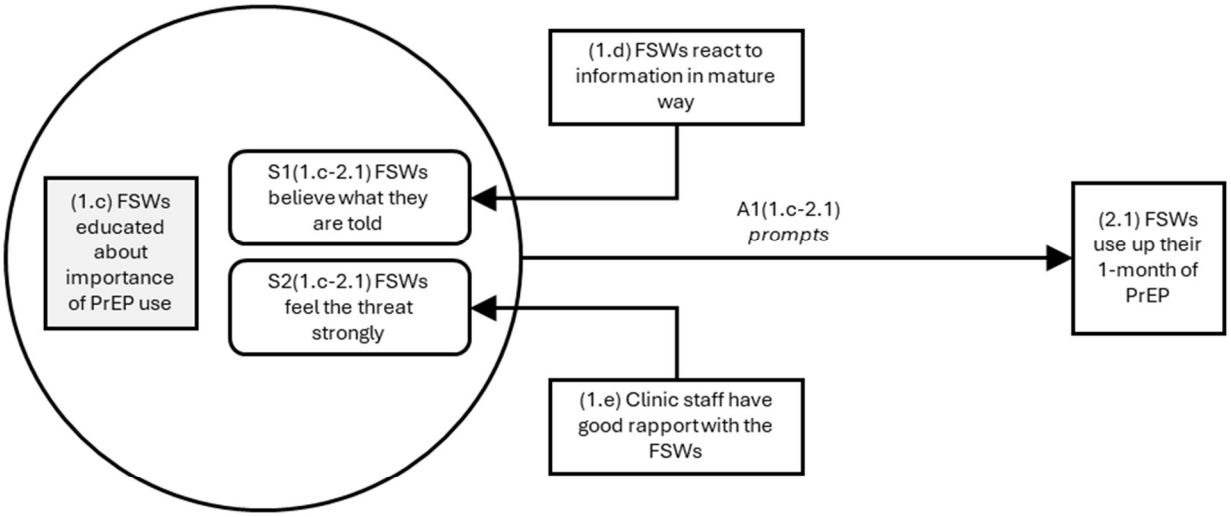
Principles:	P1/A1(1.a-2.1) 'People are more likely to initiate a behaviour when they clearly understand why it works and what benefit it brings'
	P2/ A1(1.a-2.1): 'People are more likely to complete a behaviour correctly when they have been instructed on how to complete it correctly'
Underlying Conditions:	U1(1.a-2.1): 'Service continuity during COVID-19, making access feasible and allowing services to remain open'
	U2(1.a-2.1): 'HIV and AIDS prevention and protection health information widely disseminated via many easily accessible mediums – posters, fact sheets, and brief counselling'

Figure 3. A Causal Step from a Diagram for an Enriched Theory of Change, detailing 1.a →2.1 from Figure 1.



Principles:	P1/A1(1.b-2.1): 'People are motivated to act when services are easy to access and convenient'
	P2/A1(1.b-2.1): 'People are motivated to take protective action when they perceive a serious personal heal risk and believe the action is effective and under their control'
Underlying Conditions:	U1(1.b-2.1): 'PrEP provided through convenient, familiar service points'
	U2(1.b-2.1): 'Social acceptance improved over time on public health interventions, especially those geared towards sexually transmitted infections'

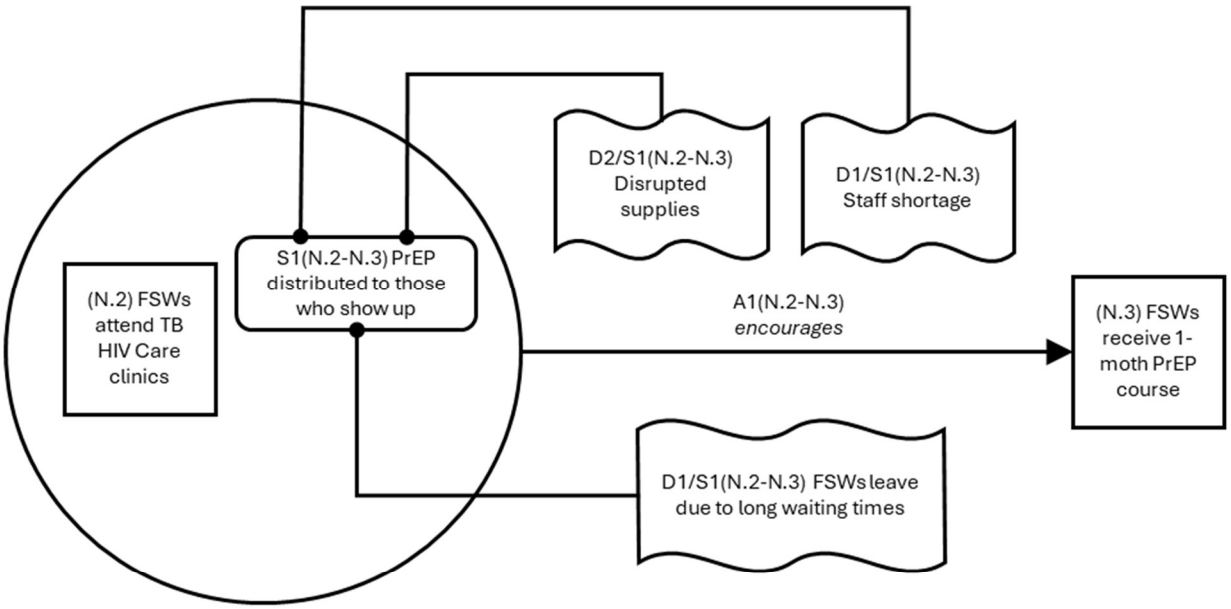
Figure 4. A Causal Step from a Diagram for an Enriched Theory of Change, detailing 1.b →2.1 from Figure 1.



Principles: P1/A1(1.c-2.1) 'People are more likely to do things they see as important and that their attention is focused on'
P2/A1(1.c-2.1): 'Timely reminders tend to reduce forgetfulness/procrastination when access is easy and non-judgmental'

Underlying Conditions: U1(1.c-2.1): 'PrEP delivery system embedded in services FSW already use'
U2(1.c-2.1): 'Proactive follow-up and friendly reminders are common in public health services in the country'

Figure 5. A Causal Step from a Diagram for an Enriched Theory of Change, detailing 1.c →2.1.



Principles: P1/A1(N.2-N.3) 'People tend to take seriously interventions promised to help prevent serious health harms if convenient'
P2/A1(N.2-N.3): 'People are more likely to return for free services, especially if they deem them to have a direct personal health benefit.'

Underlying Conditions: U1(N.2-N.3): 'Reliable stock and quick dispensing, especially for HIV and AIDS prevention and protection, is a routine service provision in many public health facilities in South Africa'
U2(N.2-N.3): HIV prevention and treatment services remained "essential" in South Africa during the COVID pandemic' (Rao et al, 2022, p. 6).

Figure 6. A Causal Step from a Diagram for an Enriched Theory of Change, detailing N.2 →N.3 from Figure 1.

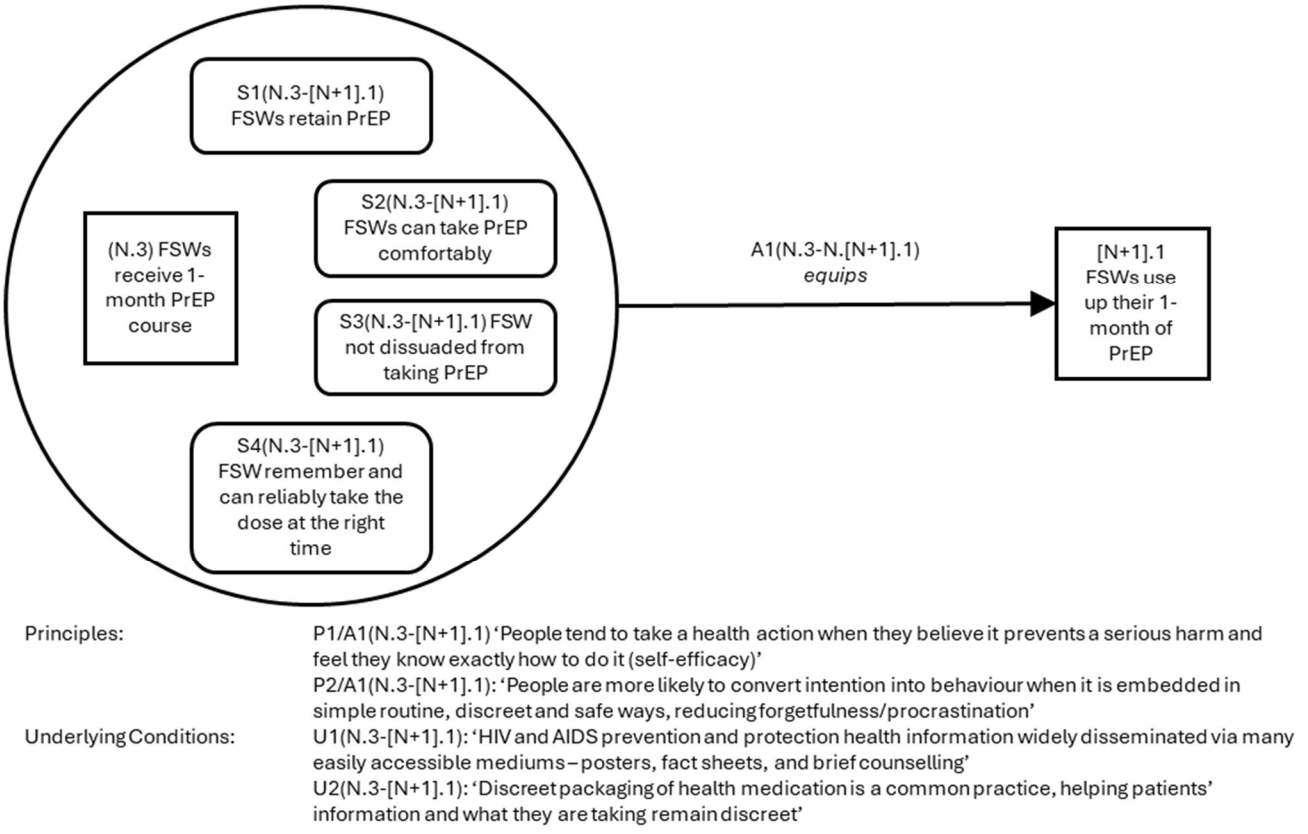


Figure 7. A Causal Step from a Diagram for an Enriched Theory of Change, detailing N.3→[N+1].1 from Figure 1.

These enriched ToCs can become large and complex. But the extra details they include say a lot about what will be needed for the policy to work in your context. Using them, you can compare whether the identified support factors, safeguards, tendency principles and underlying system arrangements are likely to be in place in your setting. Thinking in this way can also help identify plausible derailers and detractors that may arise, as well as potential safeguards against them that may already be in place – or that you may be able to implement.

Note that any ToC developed will initially be tentative and preliminary, subject to revision in light of the evidence you assemble about the local conditions (as they are or will be made to be) that will determine the effectiveness of the policy there. Given this, the business of preparing a ToC and testing it against your local setting will generally involve iterative improvements.

In the following section we describe the different kinds of evidence that can be used to support this process of iterative improvement and thereby support any recommendations that follow from your enriched ToC and show how you can marshal this evidence together.

Part 2: Understanding the evidence

Your enriched ToC shows how the policy is **supposed** to work, step by step. But will it do so or, post hoc, has it done so? In the following discussion, we use the example of appraising a policy *before* you try it but the advice is readily modified to fit the post hoc appraisal. To make a sound estimate of the policy's chances of success, ideally you should test each step of the ToC using evidence. But when this is not practicable, we advise that you start with the steps that worry you most, since the strength of the whole process depends on the strength of each step. If there is strong evidence that a step will not happen the overall policy is unlikely to succeed, at least by the process envisaged in your ToC. Strong evidence that each step will occur justifies predicting that the policy is likely to succeed.

You will notice that we employ qualitative assessments, talking, say, of 'weak' or 'moderately weak' or 'strong' support. That's because, in a host of practical cases, you will not have the background information necessary to make more precise assessments. Formal methods, like Bayes' formula, can sometimes be used to generate these. But if they are going to yield precise outputs, these methods require that you input precise assessments somewhere about something else, and you will often find that you don't have sufficient knowledge to make credible assessments there either. When you do, that is all to the good. But we do recommend a rule endorsed by Aristotle (in the *Nicomachean Ethics*) and more recently stressed by the philosopher Stuart Hampshire (in many public lectures): do not bring to a subject more precision than it can bear!

We suggest that you catalogue evidence step-by-step. You should consider, for each step, all the things that must happen at appropriate times and places if that step is to be successful and then try to gather evidence about whether they will indeed occur.

What facts matter at each step?

The facts that matter to a step's success follow immediately from the discussion in Part 1. There are seven key facts identified there:

1. **The proposed cause at that step will occur.**

If it does not, the step fails. For example, in the step from Box N.2 to Box N.3, if FSWs don't attend the clinics, then they cannot receive their month of PrEP, at least in the way envisaged by the policy. Evidence that the cause at a step is likely to occur provides some support for the success of that step. Still, this evidence is weak unless combined with evidence that the expected activity is likely to occur or that necessary support factors are likely to be in place.

2. **The proposed effect at that step will occur.**

If it doesn't, the step fails. Failures may be because the cause at that step will not be accompanied by some needed support factors, or some unguarded detractors are likely to occur or the underlying system just isn't right for this to happen.

3. **Other causes sufficient to produce an effect of the size that is supposed to occur at the time it is supposed to occur are absent**.

If other causes sufficient to cause the step's effect are going to be present at the right time, the cause at that step will be redundant. In this situation, the proposed policy is not necessary to get to that point and ipso facto to get the desired outcome by the remainder of the envisaged process. In such contexts a different cause besides the one you envisage is operating and at least overdetermining the effect. This can occur even when all the factors are in place for the step to work as envisaged.

4. **The support factors that are required will occur.**

Recall, support factors are conditions needed for the step to succeed as well as envisaged, like S1(N.1-N.2), S2(N.1-N.2) and S3(N.1-N.2) for the step detailed in Figure 2. If any needed support factors are likely to be missing, then the step cannot work as envisaged.

5. **Significant detractors and derailers will be absent (or guarded against).**

If derailers or detractors exist without safeguards, the step will fail or produce a smaller effect than hoped for (which may matter to whether the next step can succeed).

6. **The activity expected at that step will obtain, start-to-finish, and the related principles are likely to operate.**

The cause only works as intended if the requisite activity happens. So, evidence that it can do so in your setting matters, and even better is a good reason to think that it is likely to do so.

7. **The underlying system is right to allow the activities that are to occur at that step to occur, and their associated principles to operate.**

As we explained in Part 1, the underlying system in your setting must allow the activities and principles that are to operate at a step for the step to occur. If it doesn't, the step cannot work as envisaged.

Marshalling and evaluating evidence using evidence-role maps

Each of our seven key facts matter for the success of each step in your ToC, and the success of each step matters for the success of the whole process. Hence, knowing, as best you can, which of these facts will obtain and taking account of the role each plays in the process is all-important for evaluating the support you have for – or against – policy success. That means that you will need to estimate at least roughly how much support you have for each of these facts.

That some facts will obtain can just be taken for granted, given widely acknowledged background knowledge about the setting. Many will need evidence. That's a lot of evidence to keep track of and understand the significance of. To organise your evidence, we propose that for each step in your ToC, you have a **separate page** where you collect together the evidence you have about that step, dividing it into **seven columns**, each collecting evidence about one of our seven key facts.

To estimate the degree of support a body of evidence provides for one of these facts you will need to take into account that there are two sources of weakness for pieces of evidence: uncertainty about the **facts offered as evidence** and uncertainty about the **relevance of these facts**.

For instance, consider Box 2.1 (FSWs use up their 1-month of PrEP). Suppose you offer as evidence for this the claim that 'FSWs in the setting are competent users of PrEP'. This is a factual claim which is meant to make a difference to whether FSWs use up their dose. It would make more of a difference to your confidence in FSWs using up their dose if it were more certain, less if it were more uncertain. But why should you accept that it can make a difference? What relevance does being competent users have to whether FSWs use up their dose? To take it as evidence, the factual claim must be matched with a relevance claim, something like 'Competent users of PrEP are likelier to complete the courses they're given'. It is the combination of these two claims that provides evidence for 2.1.

Notice that just as with factual claims, there's room for doubt about relevance claims and room to provide evidence for and against them. One might, for example, doubt that 'Competent users of PrEP are likelier to complete the courses they're administered' because you discover that they mistrust the drug (a factual claim) and you know that people who mistrust drugs often don't take them even when they know how to (a relevance claim).

Given these two different sources of uncertainty about evidence, we suggest thinking of evidence claims not as one single claim but as a matched pair: factual claims, often (but not necessarily) expressing pieces of empirical evidence, coupled to relevance claims which explain how these factual claims bear on the conclusion to be established.

Going even further, this same approach can also be applied to any claims offered as evidence for factual and relevance claims themselves and so on, until you have fully articulated the argument for your claim, locating all evidence you have with respect to it.

This process of evidencing a claim can quickly become long and complex. One way to handle this is to draw up evidence-role maps that show how claims offered as evidence for steps in your ToC relate to the things they are intended to evidence. These are, in effect, diagrammed arguments giving reasons for or against these claims.

Throughout in applying this approach you must never forget that a lack of evidence isn't evidence against. Having no idea whether a step's activity will occur isn't the same as having evidence it won't occur. It is also important to maintain neutrality throughout about the claims in question, being sure to consider what evidence there might be both for and against them.

One way to articulate an evidence-role map is to lay things out in a diagram. For an illustration of how to do this with bundles of ex ante evidence, see Figure 8, which articulates some evidence about the likelihood of the occurrence of N.2 (FSWs attend TB HIV Care clinics) from Figures 2 and 5, or Figure 9, which maps some evidence about the occurrence of the activity 'educates' from the step 1.a to 2.1 (see Figure 3).

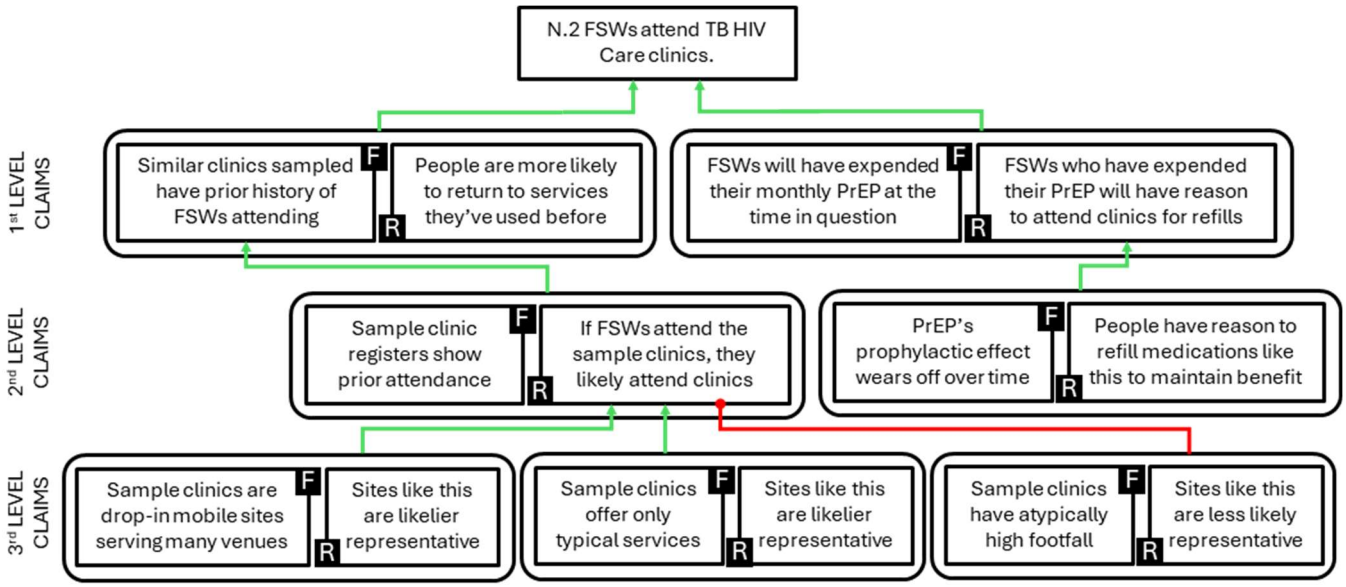


Figure 8: An evidence-role map for Box N.2 ('FSWs attend TB HIV clinics').

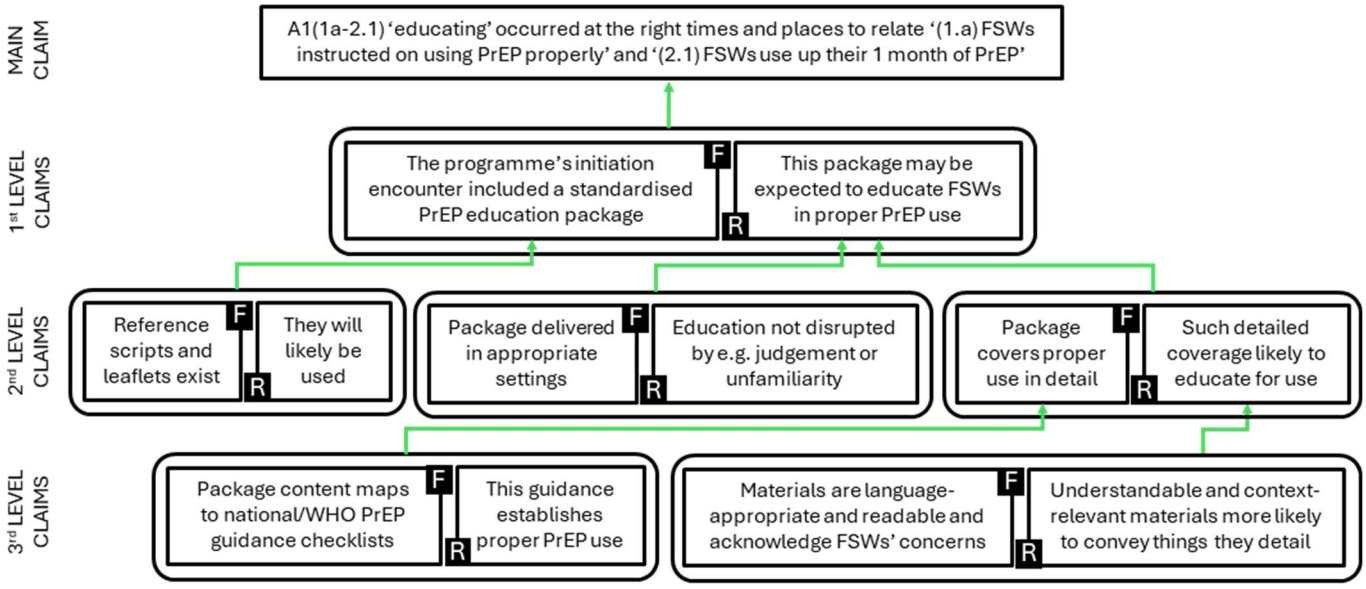


Figure 9: An evidence-role map for the activity A1(1a-2.1) 'educating' occurring between '(1.a) FSWs instructed on using PrEP properly' and '(2.1) FSWs use up their 1 month of PrEP' in step 1.a to 2.1.

In Figures 8 and 9, factual and relevance claims (annotated F and R respectively) are organised into groups which together support (or oppose) claims of fact or relevance diagrammed above them, including the 'main claim', which will be one of the seven pertinent facts for the step you're investigating.

Note that the evidence diagrammed can provide support either for or against the claim it bears on. For example, the third fact/relevance claim pair at the third level in Figure 8 argues against the first- and second-level relevance claims. We indicate the difference between positive and negative support by using a different style of connection, a line with an arrowhead at the claim end in the case of support and a line with a circle at the claim end in the case of evidence opposing the claim, as well as green for positive support and red for negative. Varying arrows or other markings can also helpfully be used to convey the strength of evidence provided by particular fact/relevance claim pairs, though, as noted, we don't expect you will generally be able to estimate strength with any precision.

Organising your evidence this way allows you to see how a piece of evidence contributes to the overall claim or how weaknesses in a piece of evidence tell against it. You need only track the connections up to see which claims are affected and hence check the impact on the overall support for the top claim. This may seem obvious in simple cases like this, but in more complex cases involving more complex bodies of evidence, being clear about just how your evidence bears on your claims can be very helpful.

Assessing chances of success

Recall that we do not expect that, in general, you will be able to do this with any quantitative precision. But understanding your evidence and just what it is doing for you should make your qualitative assessments far better grounded and accurate.

We suggest you begin by reviewing the strength of support for each step since this helps you make a sensible judgement about the success of the whole process and what design changes might improve matters.

For example, if, for a step, you have strong evidence that the cause will occur and that the associated principle often operates in settings like yours, this generally provides what we would label 'medium support' that the step will occur when you have no evidence to the contrary. If at the same time, though, you have good evidence that a needed support factor is likely to be missing, this weakens the case substantially, suggesting that things won't work as you hope. In this kind of case, it can be worth some effort to try to get that support factor into place.

Necessarily your assessments in these cases will involve judgement. Our aim is to help you improve that judgment by first, laying out, in the ways we suggest, which of our seven facts each piece of evidence is supposed to support and how it does so, and second, by understanding the role that each of these facts plays in ensuring a step's success. The use of evidence-role maps in the way we highlight should then enable both better planning and better prediction.

Some helpful rules of thumb

Below are some simple rules of thumb for assessing what the evidence indicates about how likely a step is to succeed. Note that these judgments are always relative to a given ToC, so you should bear in mind that your ToC itself may be faulty (and be prepared to revise it if you find one that better fits what happens in your setting).

In what follows, we mean by 'good' evidence, fact/relevance pairs that supply a strong reason for belief. We assume for these purposes that your ToC describes your context accurately and doesn't miss anything important.

For any given ToC:

- Good evidence that the cause will occur (fact 1) and that the envisaged activity will occur in accord with appropriate principles (fact 6) together amount to strong evidence that the step will occur.

- Good evidence that the effect will occur (fact 2) and that no other sufficient cause of this effect will occur (fact 3) together amounts to strong evidence that the step is not redundant and so will be able to make a difference.[3]

- Good evidence that the cause will occur (fact 1) and that sufficient support factors will occur (fact 4) together amounts to strong evidence that the step will occur as envisaged, if no detractors strong enough to derail the step are likely.

- If there's good evidence that no set of factors strong enough to derail the step will occur or that they will be guarded against (fact 5), this together with good evidence that the cause will occur (fact 1) and that sufficient support factors will occur (fact 4) amounts to very strong evidence that the step will occur.

- Good evidence that any one of the seven facts necessary to a step working will not occur is, by itself, strong evidence against the step working.

- Good evidence that a cause will occur (fact 1), or its effect will occur (fact 2), or its support factors will occur without or with safeguarding against derailment (facts 4 and 5) or that its activity will occur (fact 6) amounts to weak evidence that the step will occur.

[3] This 'Holmesian inference' approach is especially useful in post hoc assessment, where evidence about the occurrence of steps' respective effects is more likely to be available.

Part 3: Better ToCs, better appraisals

What can you learn from working through the justifications for the causal claims involved in policy in this way?

Most clearly, you can gain insight into, and better express, how well-justified the claims are, given your evidence. A well-constructed ToC, step-by-step validated by well-mapped and high-quality evidence, allows you to assemble strong support for or against a causal claim. Equally, you might discover a lack of such support that might otherwise be missed, which is no less important in working towards well-evidenced and effective policy. Furthermore, you are doing all this by drawing on evidence sensitive to the context in which the policy is implemented, allowing local variations in how things work and what can work to emerge during the step-by-step validation and refinement of your ToC. This gives a basis for interrogating the likely effectiveness of both imported policies and home-grown ones.

A further benefit is that, by organising the evidence for your ToC in the way we suggest, you can gain insight into where gaps in this evidence exist. This will be wherever evidence-role maps show a reliance on claims that seem in need of further justification, or where what appeared to be a sound theory of a causal step turns out to rely on unsafe assumptions about any of our seven key facts pertinent to that step. In such cases, you can return to evidence collection armed with greater clarity on what needs to be checked and thereby more efficiently develop support for or against the policy you are considering implementing.

Of course, evidence about which of the necessary factors are already in place and which are not, and whether your local setting can reasonably easily provide those that are missing, will be invaluable for designing how to implement your policy and for predicting the chances of success. If the underlying system looks unlikely to allow for the causal pathway by which your policy is expected to work, then you must assess the policy as unlikely to work in your setting, at least in the way envisaged, and – as a planner – consider alternate pathways by which to achieve the intended results (or else pursue more systemic change). You can equally consider alternate means of pursuing your intended outcomes if the support factors, safeguards or tendency principles look unlikely to occur as needed by the ToC you envisage, or if you envisage unsafeguarded detractors or derailers. You can also – in planning implementation – take steps to arrange for needed factors to be put into place or problem detractors or derailers to be removed, as practicable. By enriching and interrogating your ToC in the way we suggest, whichever of these approaches you are taking, you should be better equipped to assess and ultimately maximise your chances of success. At least, you should be better equipped to justify your assessments in a clear, orderly and accountable way.

The enriched ToCs and evidence-role maps that we recommend may look daunting. However, such detail is necessary for making informed judgments about what is happening in the social and physical world. Note that reliable quantitative methods for processing evidence are just as onerous when written out in full, and they are onerous for just the same reason. The world is

complex, and effective attempts to assess and plan interventions in the world must grapple with this complexity. The effort put in will, we believe, help you plan wisely when deciding whether an intervention can work in your context and in planning how to implement it more effectively.

Try it and see:

We suggest if you are considering using this method, that you try it out on one step in the ToC. identifying one that you think might be the weakest link, the hardest to implement well. This will give you a taster of how useful it can be.

1. **Interrogate that step using the Seven Key Facts checklist** (page 3).
2. **Catalogue evidence** for each key fact (even rough local knowledge counts but label it).
3. For each evidence item, write two linked claims:
    - **Factual claim:** what the evidence says about the world here
    - **Relevance claim:** why that fact matters to this step in the ToC.
4. **Judge support** for each key fact (strong/medium/weak), noting:
    - What would most improve confidence
    - What design changes could reduce risk.
5. **Revise the ToC** (and implementation plan) accordingly.