Evidence-Informed Policymaking

Practical tips and inspiring case studies for more effective decision-making
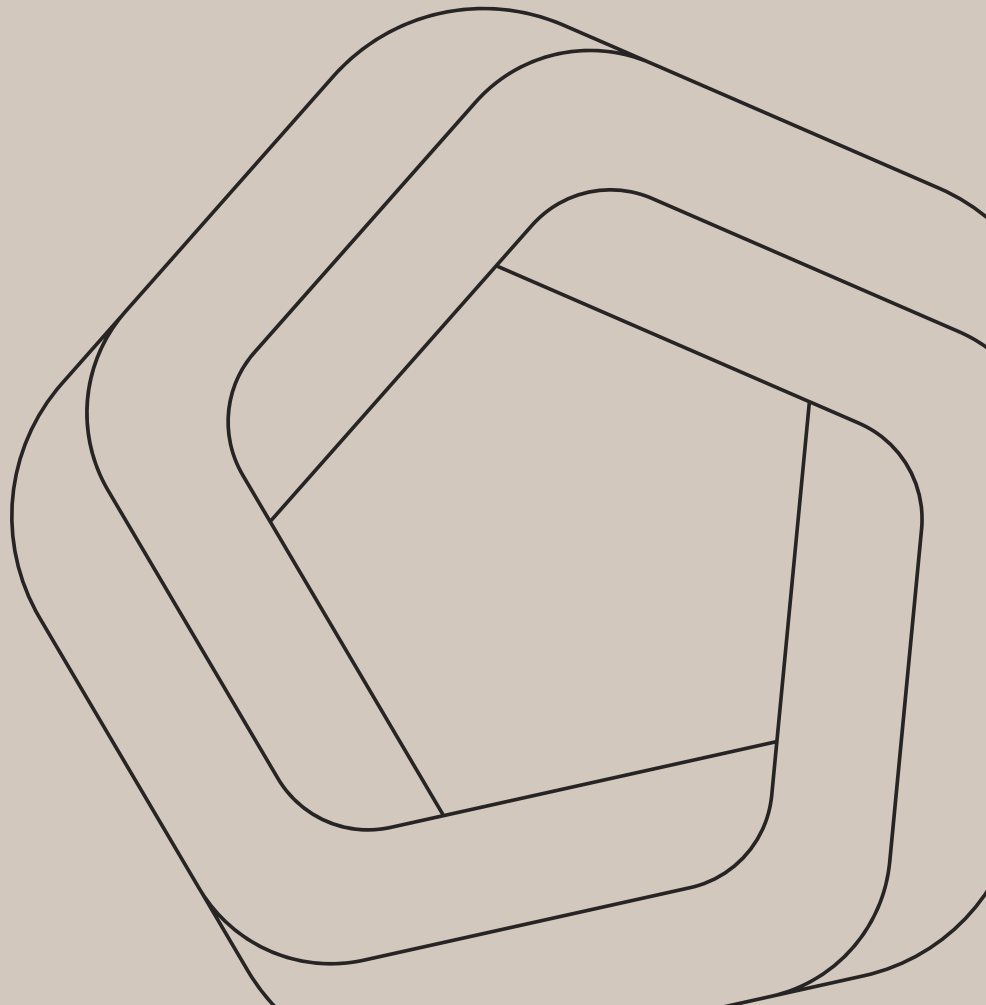
Policy Brief

By Elke Himmelsbach

Dr Sophia McDonnell

2025

Contents

Summary

Evidence-Informed Policymaking – Why and How It Works

This practical guide demonstrates how both scientific and exper ence-based evidence can be systematically integrated into all stages of the policy cycle – from identifying and diagnosing problems to developing, implementing and evaluating solutions. Drawing on inspiring case studies from Germany, Europe, and Australia, the guide is aimed at decision-makers seeking more effective, evidence-driven and broadly accepted public policy.

Relevance for Policy Practice:

Political decisions are often made under time pressure and based on routines, intuition or political constraints. Yet the message is clear: using more and better evidence leads to more efficient measures, stronger public support, and greater impact – particularly in today's environment of complex and polarised policy issues.

Key Themes and Takeaways :

- **What constitutes "good evidence"?**

Differentiation between scientific (quantitative/qualitative) a d experiential evidence. Both are valuable – ideally in combination.

- **Evidence across the policy cycle:**

  » Agenda setting: Prioritising based on urgency, data gaps and political feasibility
  » Diagnosis: Analysing root causes using behavioural science insights
  » Design: Developing, adapting and testing effective policy options
  » Implementation: Addressing resistance, engaging stakeholders, ensuring scalability
  » Evaluation: Measuring impact, cost-effectiveness, long-term outcomes, and learning loops

- **Real-world case studies:**

Examples include combatting violence against women in Australia, increasing bicycle helmet use in the Netherlands, improving teacher training in Germany, and shaping bottle recycling policy in France. These and other cases demonstrate how evidence not only improves policy effectiveness but also helps build consensus and resolve societal conflicts

Practical Recommendations for Policymakers:

- Sound evidence is trustworthy, objective, relevant, up-to-date, and methodologically robust.
- Successful evidence-informed policy requires: low-threshold tools, clear processes, interdisciplinary collaboration, and data literacy.
- Even under high political pressure, evidence enables more legitimate, transparent decisions and helps build public trust.

Conclusion:

This guide provides actionable impulses for integrating evidence into everyday policymaking. It does not just argue why evidence matters – it shows how to make it work. A practical toolbox for all who aim to increase public impact and restore trust in government.

Why evidence-informed decision-making?

There is a growing need for more evidence-informed policy-making. When political responses, laws and the implementation of policy programmes do not match the perceived reality on the ground, disappointment and dissatisfaction among citizens and companies increases (Ludewig in NKR 2019).

Research on evidence-based policy-making shows (Sager et al 2021) that integrating more evidence into political decisions improves the effectiveness and efficiency of initiatives and can help regain citizens' trust. However, a recent survey by the German National Academy of Sciences Leopoldina (Riphahn and Schnitzer 2022) found that political decision-makers in the Bundestag do not use as much scientific evidence as they would like, and as the OECD (2021) and other national audits (NKR 2019) suggest.

There are many obstacles to the use of evidence:

- **Habit or tradition** of using personal expertise instead of data (Hadorn et al 2022); stakeholder engagement of social partners and experts is well established, but there is no equivalent for other audiences such as involving citizens (OECD 2021)

- **Lack of ease of access,** i.e. evidence provision is dominated by long-standing institutionalised knowledge transfers with outdated procedures (NKR 2019, Hadorn et al 2022)

- **Lack of good communication** of scientific evidence between providers and decision-makers, e.g. lack of short and comprehensible summaries and visualisations, transparent methodologies, unbiased approaches, clear differentiation of recommendations that fit into the context of the decision-making process (Seidel et al 2021)

- **Lack of time,** i.e. political decision-makers are often pressed by faster media cycles to quickly commit to seemingly obvious solutions that may not even be fully developed yet, and this leads to an unfortunate habit of defining solutions too quickly and in too much detail (NKR 2019)

Political decision-makers have access to evidence in principle, but many cannot make use of it appropriately. In practice, political decision-makers use values, personal experience, and political necessity as well as evidence to inform decisions, so they apply a limited or 'bounded rationality'. Despite the rhetoric around the importance of evidence, the proper use of evidence for policy remains challenging and somewhat elusive.

Therefore, there is a need for better toolboxes to make it easy and feasible for political decision-makers to use evidence in their daily work in Germany. As a first step, this paper aims to provide inspiration on how to overcome these barriers and how to use evidence in various stages of the policy cycle.

The benefits of using evidence:

Optimising resource allocation

Enhancing public service delivery

Achieving better outcomes for society

What is good evidence?

When we talk about evidence, we recommend keeping the following two types of evidence in mind (WHO 2021):

Scientific or research evidence	Experiential evidence
Is produced systematically through a formal, standardised research process following certain methodological standards.	Constitutes more informal knowledge based on the opinions and experiences of people who work or live in the area being analysed.
It can include primary research, synthesis of existing research (secondary research), and other products and guidelines based on evidence such as guidelines (tertiary research). Scientific evidence can be numerical (quantitative) or descriptive (qualitative), using a range of methodologies.	This type of evidence is often expressed in the media, expert reports, personal anecdotes, qualitative interviews, group discussions or deliberative dialogues.

Both types of evidence can be valuable and answer different research questions. Their relationship is complementary. For example, experiential research can highlight an area to focus scientific research on, or it can validate or question the scientific research already carried out.

Typical research questions that can be answered by each respective evidence type (Superu 2018):

Scientific or research evidence	Experiential evidence
☐ Does the evidence validate, or accurately describe, your initial hypothesis, the target population and/or the drivers for change?	☐ How does the policy approach reflect the experiences and knowledge of the people it aims to serve?
☐ Is there evidence for potential programmes, practices or policies being effective in addressing the issue you want to resolve?	☐ What can a target group's earlier experiences tell you about the acceptability, importance and appropriateness of the policy or programme to them?
☐ Are there well-designed studies or evaluations available to support or validate the success factors or outcomes you are seeking?	☐ What can the experiences and knowledge of stakeholders tell you about what has previously worked or not worked with the target group?
☐ What positive, negative or unintended effects has a programme, practice or policy had on changing behaviour or outcomes?	☐ What common goals do the stakeholders and target group have around the issue that the policy addresses?
☐ Is there any implementation guidance available? What does the evidence say about the resources, processes and capacity needed to successfully implement a (change in) policy?	☐ What does the evidence tell you about how well matched the experiences and aspirations of the target group are to the goals of the policy?

☐	What are the characteristics of the people the policy will serve? How and to what extent are the people affected by the issue, problem or opportunity?	☐	What experiences have been captured from evaluations or reports from similar initiatives? How could they be used to strengthen your approach?

How can you judge good evidence?

When deciding which evidence to base decisions on, it is important to assess the robustness and quality of the available evidence. There are numerous guidelines from research associations and institutions that provide a detailed overview of quality standards for the production of evidence.
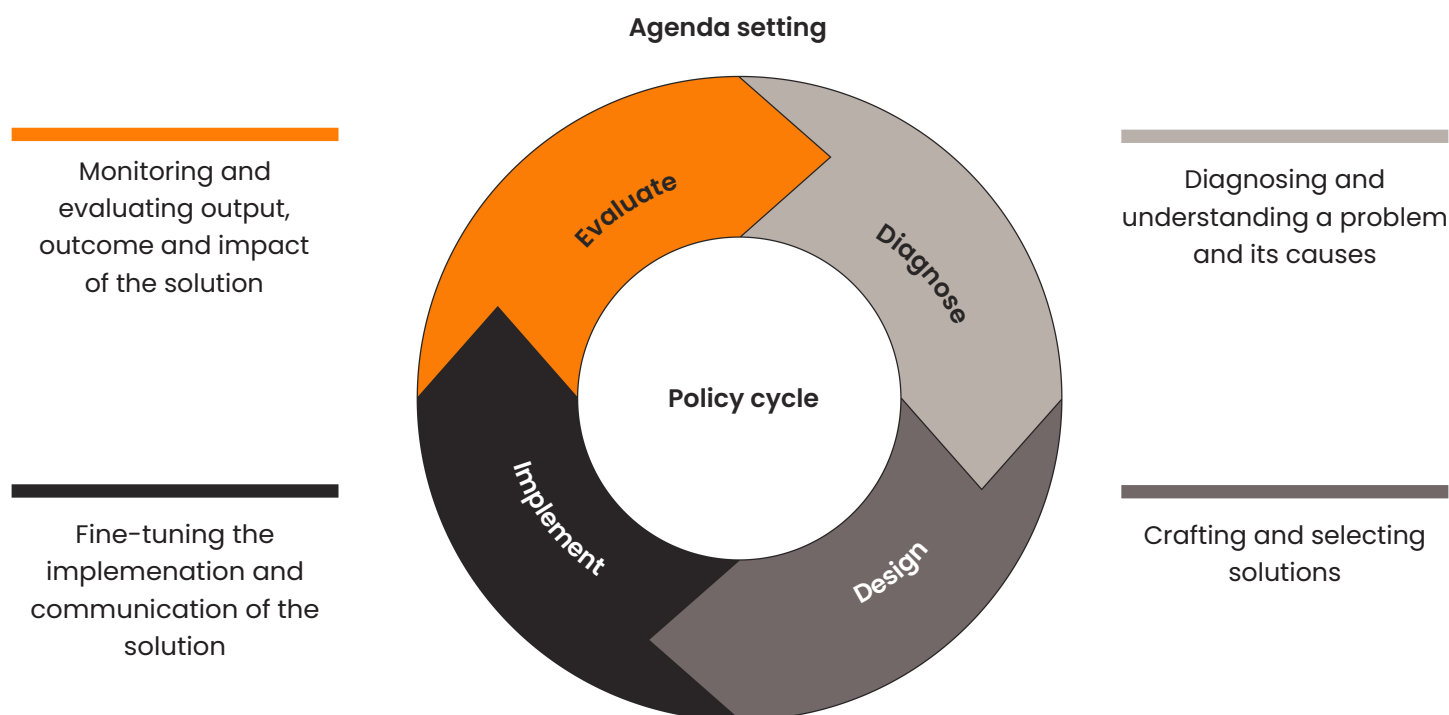
In the German market, the 'Gesellschaft für Evaluation e.V.' (DeGEval) provides a good overview of the standards for evaluations and 'Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.' ADM e.V. offers a useful checklist for clients commissioning surveys. In practice, however, it is more useful to be guided by concise and clear frameworks highlighting key dimensions of good evidence, such as the following framework from evidence-based policy practice in the UK (Houses of Parliament 2017, Nesta 2022).

When trying to assess evidence from a variety of sources and methods, it is helpful to ensure that it fulfils the following criteria:

Authoritative	It is important to assess where the information comes from and who created it. Are the qualifications of the authors credible and is their approach transparent?
Objective	To understand the aim behind any research and the possible biases that may occur, you should consider the purpose of the study and the likely motivations behind it. In this context, it is often useful to check who funded the research.
Relevant	Not all research is useful to your question or situation. You should therefore check whether the context in which the research took place is comparable to your own, whether all relevant perspectives were included, and which target group the research was originally intended for.
Timely	If you are basing your decision on evidence, it is important that this evidence is not out of date and is still valid. You should therefore check whether the information you have is up to date and whether anything important has happened in the meantime that could affect the results.
Accurate	To assess the quality of a method, you should ensure that the methodology is described in detail and is appropriate. It is also helpful to check if similar studies have been conducted or whether the study has been repeated. As a rule, peer review or validation by other experts in the field is a good sign of strong evidence.

The use of evidence in the policy cycle

Evidence can make you and your policies more effective and powerful. This chapter provides you with ideas and inspiring case studies that showcase how evidence can support your political decision-making at different stages of the policy cycle.

Agenda setting

Monitoring and evaluating output, outcome and impact of the solution

Diagnosing and understanding a problem and its causes

Evaluate

Diagnose

Policy cycle

Implement

Design

Fine-tuning the implemenation and communication of the solution

Crafting and selecting solutions

The development of a policy measure usually begins with an agenda-setting process in which the topics that are considered relevant are selected. The diagnose phase focuses on understanding and describing the problem and its causes. This is followed by the design phase, in which solutions are developed and selected for the next step, the implementation phase.

During the implementation phase, a number of decisions are usually made to optimise the communication and processes. Last but not least, the implemented solution should be monitored and evaluated to determine whether it has solved the original problem and achieved the desired goals.

The findings from the evaluation are then incorporated into the next agenda-setting, not only for the same problem (if it has not been solved), but also for similar problems to which the findings can be transferred.

DIAGNOSE: Understanding a problem and its causes

The policy cycle begins with the identification and formulation of a problem that requires attention. Ideally, this step should be preceded or accompanied by a formal prioritisation process that considers the answers to the following questions:

Agenda setting **Should we pay attention to this problem?**	**How big is the problem?** > What are suitable indicators for measurement? **Is it getting worse or is it bigger here than elsewhere?** > What are comparable benchmarks? **How long will a priority last in a constantly changing world?** > When to collect which types of evidence? > When is evidence outdated?

The answers to the above questions will lead to the description of the problem, which can look different depending on the perspective. The understanding of the problem should therefore be underpinned by research evidence, the decision-making context, ethical considerations and discussions with key stakeholders.

Context & Diagnosis **How can we describe the problem and its causes?**	**What are the relevant context scenarios and life situations in which the problem occurs?** **How do different people describe and experience the problem and its causes?** > Who are the relevant stakeholders? **How can causes of a problem be identified?** > ...especially when they are not obvious

The most difficult questions to answer in the diagnose stage are the latter ones, which is why we present potential solutions based on two case studies below.

The first example illustrates the temporal challenges in the agenda-setting phase based on the election campaign decisions made by political parties. It is based on the experience of Verian in Germany.

How long will a priority last in a constantly changing world? When to collect what type of evidence? When is evidence outdated?

Before elections, political parties have to make decisions about the candidates and the issues on which they want to base their campaign.

The party committees usually opt for an evidence-based briefing before making their decisions and briefing the communications agency.

The biggest challenge is the timing: when to collect what kind of evidence?

Political moods can be very volatile. An early start allows for an early briefing with enough time to develop a good campaign. But it also carries the risk that many things can happen that change the perception and the impact of a campaign. **The solution is to separate strategic and tactical objectives.**

Typical strategic research questions are:

- Who are the accessible voter segments among eligible voters?

- Which topics and narratives mobilise potential voters and which topics demobilise political opponents?

- Where is our party vulnerable?

Typical tactical research questions are:

- What is the central campaign message in the final election camp ign phase that has the best potential to maximise votes?

- How to deal with postal voters who might vote up to six weeks before the election date?

Evidence for strategic questions is usually collected 6 to 12 months before election day. Evidence for tactical questions is gathered about 4 to 12 weeks before election day. **However, if unexpected events occur** (e.g. a new conflict or a new opposing candidate) **it is advisable to update the strategy based on new evidence collected after such significant events.**

Key learning for the diagnose phase, regardless of your policy area:

1. Divide the issues into strategic/long-term vs. tactical/short-term.
2. Choose timely evidence when prioritising between different issues, people, etc.
3. Check whether the existing evidence is overtaken by important events or trends.

The second example illustrates the importance of using holistic diagnostic tools that incorporate multiple perspectives to identify the root causes of long-term social problems. Only then is it possible to identify levers for changing trends. It is based on the experience of Verian in Australia.

How can causes of a problem be identified, especially when they are not obvious?

Violence against women has a long tradition in Australia. Two out of five Australian women have experienced violence by a man after the age of 15. Every week, one or two women were killed by a current or former partner. The consequences of violence against women have cost the Australian economy 13.6 billion AUD annually. No wonder, the Australian government made reducing violence against women a national public health priority in 2015.

While there was extensive existing research on the subject, the vast majority had focussed on identifying the scale of the problem and the rational, conscious attitudes that can affect it.

However, there was a lack of understanding of the core of the problem and its starting point, and therefore no idea of what could be done to reduce the violence.

New research revealed that social norms and unconscious beliefs among adults and young people were promoting gender inequality, disrespect and aggressive behaviour. For example, there was a tendency towards victim blaming, when examples of aggressive or disrespectful behaviour between men and women were brought up ('she must have done something'); as well as empathising with males ('it's very hard being a young boy, he probably just likes her') and trivialising behaviour (it's not that bad … it's not like she ended up in hospital'). Due to the unconscious nature of these engrained heuristics, many Australians become reconciled to their role in perpetuating the situation and easily dismiss responsibility.

This is a result of the perceived high cost of exerting influe ce and, as a result, many are quick to accept not to get involved when disrespectful behaviour between young men and women occurs. It is a common norm to view this experience as an important part of youth development. Infl encers need to feel more confidence and self efficacy in order to respond appropriately.

The result of the initial evidence base was the development of a long-term national behaviour change campaign aimed at challenging and changing the engrained attitudes and beliefs at a societal level over time. The 'Stop it at the Start' campaign is a long-term investment, commensurate with the scale and complexity of the issue – because changing deep-rooted societal attitudes does not happen overnight. It is a primary prevention campaign. Since its launch, it has been independently evaluated and is recognised as the most successful campaign ever launched by the Australian Government.

It has led to 'new conversations' about the links between disrespect and violence against women, a significant reduction in the unconscious heuristics that unintentionally perpetuate the problem, and a significant reduction in violence-supportive attitudes among young people. And ultimately, the average number of Australian women killed by a current or former partner has decreased significant y from one or two every week (2015) to one every 11 days (2022).

What kind of evidence helped to uncover the cause of the problem?

Qualitative research to deconstruct the crux of the issue: to understand why, when most Australians condemned violence against women and children, the problem persisted in a pervasive way in society. Advanced qualitative techniques were utilised in combination with Verian's proprietary Behaviour Change Framework.

Verian collected the perspectives of many different groups in different settings: Focus groups were conducted with mothers, fathers, parents together, older and younger siblings, grandparents and non-parent influencers. In-depth-interviews took place with more vulnerable target groups such as people with disabilities, members of the LGBTQIA community, and survivors of domestic violence.

Key learning for the diagnose phase, regardless of your policy area:

1. First check the available data and identify important gaps of evidence. Gaps can be missing perspectives of those involved, or missing descriptions of the causes and influences of the problem

2. Apply holistic thinking, in particular when discussing a long-term problem where there are no obvious causes and solutions. Integrate insights and methodologies from the behavioural sciences to understand why real behaviour often differs from the intended behaviour.

DESIGN: Crafting and selecting solutions

Once the problem and its causes have been identified, the next tep is to develop and select the best solution for this problem.

Depending on the nature of the problem and the degree of novelty, we may need an exploratory crafting phase for new solutions or a literature review for identifying best practice solutions. In most cases, there will be a variety of possible solutions which usually require evidence to show which solution offers the highest impact and/or the best cost-benefit ratio. The following questions are typical of decision-making in this phase of the policy cycle:

Design & Decision How can we identify the best solution to the problem?	**How can we generate solutions for addressing the problem?** > Can we adapt something that worked elsewhere? > How to craft new solutions? **How to identify effective or cost-effective solutions?** > How to test and assess the options? > Does one option achieve more than another with the same investment? > How to apply a good theory of change or logic model? **How to ensure stakeholder buy-in?** > How to appropriately include stakeholders or use deliberative methods and collaborate/co-create?

The first example below illustrates the challenges in the desig phase when searching for potential solutions. It is based on the experience of the Swedish National Council for Crime Prevention and Malmö University in Sweden, who accompanied the design phase with a process evaluation published in 2020.

How can we generate solutions for addressing the problem? How to adapt something that worked elsewhere?

Serious violent crime is one of the major societal problems in Sweden. The main source of this violence appears to be within and between criminal groups and is related to fire rms. Swedish research on what works to reduce serious violence is limited, and more knowledge is needed to provide the police and other crime prevention actors with the tools they need to reverse the trend.

To identify potential strategies for combating lethal violence, the Swedish National Council for Crime Prevention (Brå) looked at regions around the world that had previously tackled lethal violence. They found that a strategy called 'Group Violence Interventions' (GVI) had been successfully applied in Boston and other American cities in the 1990s.

To learn more about the requirements and practical implementation of this GVI strategy, representatives from Brå conducted a study visit to the United States and consulted with Professor David M. Kennedy, who led the Boston Gun Project in the 1990s and had become the Director of the National Network for Safe Communities in the United States.

The GVI policy that worked so well in Boston was developed by an interdisciplinary working group in a step-by-step process that helped the team identify and effectively solve problems that they may encounter (problem-solving model approach). The working group consisted of frontline practitioners from a variety of agencies as well as academics. Together, they combined qualitative and quantitative research with the experience of police officers and streetworkers.

The GVI approach sought to deter gang violence by convincing gang members that violence and gun use would have consequences, causing them to choose to change their behaviour. A key element of the strategy was to deliver a direct and explicit 'deterrence message' to a relatively small target audience, which spelled out the response to certain behaviours. Boston Police Department made engaging in gang-related violence more costly for those involved. Communicating what had happened to others in the target audience was expected to deter further gang-related violence in the city.

After the successful implementation of GVI in Boston, the strategy has been implemented mainly in American cities, but also in a few European cities such as Glasgow and London. The conditions for being able to implement GVI in Sweden have been questioned with reference to the fact that the differences between the United States and Sweden are too great.

Therefore, it was decided that a pilot project would be set up in one Swedish city. This 'real world' trial was accompanied by a process and impact evaluation to understand the elements required to adapt such a policy to different context/areas. Malmö municipality became a pilot project for GVI in Sweden for two years. The American team was consulted for its knowledge based on its role in Boston's previous GVI work. The impact evaluation of this trial, published in 2021, showed that the number of shootings in Malmö decreased during the period of the pilot project. However, it is difficult to ascert in statistically what factors are behind the reduction. Nevertheless, the results were promising enough for more Swedish cities to adopt and adapt this policy in their areas.

Key learning for design phase, regardless of your policy area.

1. Look at what worked elsewhere but be aware that adjustments may need to be made.

2. Make sure you understand the underlying impact model of the policy that can be used regardless of the local requirements.

3. Assemble a team of academics and practitioners to develop and adapt policy solutions to increase the likelihood of them working in reality.

Since it is not always possible or economical to conduct real-world trials to test the effectiveness of one or more solutions, the next example illustrates how such a test can be carried out faster and more cost-effectively as laboratory experiments, that compare several solutions or variations of solutions. It is based on the experience of the Behavioural Practice at Verian in the United Kingdom.

How to identify effective solutions? How to simulate and test policy measures in the design phase?

According to Ofcom, the UK communications regulator, most internet users believe the benefits of going online outweigh the risks. However, seven out of ten users report having seen or experienced something potentially harmful online in the last three months. Online platforms commonly use safety measures such as alert messages, reporting mechanisms, parental controls and terms and conditions to safeguard their users.

However, there was limited research in the public domain about their effectiveness. At the same time, there is growing behavioural insight evidence that relatively small changes to the design of the online environment, which could include safety features, can shape user behaviour – for better and for worse.

To learn more about the impact of specific alert messages and c ntent-reporting mechanisms on video-sharing platforms (VSPs), a randomised control trial (RCT) was embedded in an online survey that simulated a natural VSP (such as YouTube, Vimeo or TikTok). This laboratory experiment allowed the observation of typical behaviour when scrolling through a social media feed. Each participant interacted with a mock-up of a VSP interface and had the opportunity to view a series of short videos: some containing neutral content, and some containing legal but potentially harmful content. Participants were randomly allocated either to a control group or to a treatment group. The control group was shown an interface without an alert message. Each of the four treatment groups received a specific type of warning messages.

The findings suggest that alert messages prompted participants o become more engaged in making decisions about watching the video content they were presented with, but only the high level descriptive social proof alert message "This video contains material that other viewers on this platform have reported as being sensitive" had a statistically significant impact on whether participants skip ed legal but potentially harmful content.
This form of alert message gave those participants who did not wish to view the content the information to enable them to decide to skip that content. Equally, those participants who chose to view the content after being exposed to an alert message were better placed to understand the risk of harm: they were either comfortable with the content or were alive to the fact that they might find the content problematic and so were 'primed' to skip away from it if necessary.

The most effective solution in this online experiment also showed no negative effects, as it did not increase the likelihood of skipping neutral content and it was not considered to be annoying.
For Ofcom, this was the first experience of an online RCT using a mocked-up user interface to simulate a typical (and often unconscious) series of decision-making scenarios. The experience showed that, even if the results are not observed in a real-world environment, a good simulation can provide a good balance between providing high-quality evidence within a short time and a feasible budget. The experimental results are more realistic than a simple survey with hypothetical questions about behaviour.

A larger sample size makes it possible to test and compare several treatments, and the laboratory design allows all other extraneous influences to be controlled, in order to i olate the effects of the treatments (which is not possible in a real-world design). Ofcom would therefore like to continue using this methodological approach.

Key learning for design phase, regardless of your policy area.

1. When testing various concepts/versions of policies, consider RCTs (randomised controlled trials) that simulate the decision-making situation with realistic context and choice architecture for observing realistic behaviour.

2. Include all potentially influencing factors in the research to control them and to better understand the effect of the treatment.

3. For measuring long-term impact of the treatment on behaviour change, it is necessary to implement follow-up waves for observing repeated behaviour or sustained attitudinal changes.

The next example illustrates the challenges to ensure stakeholder buy-in in the design phase when having to select a solution that is highly disputed among different parties. It is based on the experience of Verian in France.

How to ensure stakeholder buy-in? What to do if there is strong political pressure on selected policy options?

The French government has set itself the ambitious target of increasing the recycling rate of plastic bottles from 61% in 2021 to 90% by 2029. In addition, the country aims to halve the consumption of plastic bottles by 2030. To achieve the new targets, the Minister of Ecological Transition planned to introduce a new deposit system for recycling plastic bottles.

This planned solution was highly controversial. According to its proponents, the beverage industry and environmental organisations, the deposit system would be an important lever in achieving the recycling targets set by the EU. Other countries, such as Germany, show that such a system can be effective in achieving recycling targets.

Opponents of this solution are the local authorities, which have already invested heavily in the current recycling system. A switch to a deposit system would deprive the local authorities of a source of income. There was also concern that the most valuable waste would be returned to manufacturers for recycling, but that local authorities would be left with the worthless waste that is not returned and is most difficult to recycle.
It was therefore a happy coincidence that a law passed in 2020 obliged the Minister to base the political decision on solid evidence.

ADEME, the French Environment and Ecological Transition Agency, commissioned a research consortium (Verian, Ifop and EY) to collect evidence to evaluate the concept of a new deposit system for beverage bottle recycling compared to the existing recycling process.
The evaluation was based on a number of data sources: European benchmarking of deposit systems, a quantitative consumer survey on purchasing and sorting behaviour, RCTs to test consumer perceptions and practices for the new concept compared to the status quo, and an economic analysis to examine the economic viability of the new concept compared to the current one.
The evaluation results published in June 2023 showed that the new deposit system would not fully achieve all set targets. Furthermore, the current system would also be able to achieve the targets, provided that all the planned resources were deployed.

In view of these conclusions, but also considering the strong resistance of local authorities to the new deposit system, the Minister of Ecological Transition decided in September 2023 to abandon the plan of a deposit system for plastic packaging. However, he announced that this would be offset by setting significant targets for local authorities to improve collection and recycling, based in particular on the strategies identified by ADEME.

Thanks to this evidence-based decision-making, the reactions of the stakeholders and the media were overwhelmingly positive: "government backs down from a 'false good idea'", "a victory for our territories, for our residents and in the fight against the climate emergency in view of the budgetary and ecological impact that this measure would have entailed".

Key learning for design phase, regardless of your policy area.

1. Topics with high political pressure require trustworthy and independent sources.

2. Ensure that all relevant stakeholders are heard, and the most important arguments are examined.

3. Evidence can help decision-makers to create acceptance for their decisions.

IMPLEMENT: Fine-tuning the implementation of a solution

Once the decision has been made as to which solution should be implemented for the problem, the next phase involves fine-tuning decisions during the implementation phase, including plans for communications. This phase is often neglected by political decision-makers, as they usually focus on the macro level of a political solution.

However, the success and effectiveness of a solution often depends on well-designed implementation and communication in practice. To ensure that the policy works as efficiently and as effectively as possible, it is important that the implementation concept is carefully scrutinised and tested, ideally with all relevant intermediaries and target groups. Typical research questions at this stage are:

Implement & Communicate Should we take any additional steps to increase the chance that the selected solution does what we intend it to do?	What could cause the measure to fail during implementation? How can a policy be fine-tuned with additional insights for a communications strategy? How to ensure participation of intermediaries and target groups over a longer implementation period? How to scale interventions that have been found to be effective during a pilot?

The first example below shows how resistance to policies and pr grammes can be dealt with through communication strategies. It is based on the experience of Verian in the Netherlands.

What could cause the measure to fail during implementation? How can a policy be fine-tuned with additional insights for a communications strategy?

The Netherlands is known as the "land of bicycles" and has the highest rate of bicycle ownership in the world. At the same time, the country has experienced a notable rise in cyclist casualties. A key factor in improving safety is the usage of bicycle helmets, but very few people in the Netherlands wear helmets during regular bike rides.

In response, the Ministry of Infrastructure and Water Management wanted to develop a comprehensive bicycle safety plan to reduce the number of serious injuries among cyclists. However, in order to be picked up by citizens who seemed predominantly opposed to helmet usage, research into what was driving attitudes, beliefs and behaviours of Dutch citizens on voluntary helmet usage for regular bike rides was required.

Only with the knowledge of key driving factors of resistance against helmet usage can effective communications and implementation measures be designed and policy measures on helmet use actually pursued by citizens. As influencing factors can be hard to pin down for a topic whe e the prevailing social norm is so strong, Verian designed a two-stage research project. Interviews with parents of young children, commuters and seniors who cycle regularly uncovered that the main barriers to helmet usage were a low-risk perception, and the belief that helmets are unnecessary and affect their image negatively.

Key barriers included social norms, stereotypes, appearance issues and inconvenience. However, the research also showed that acceptance could be increased if helmet use becomes more common, risk perception rises, design and comfort improve, and personal values like responsibility and health consciousness become more associated with helmet use.

The quantitative survey showed that a considerable sub-section of non-helmet-wearers would be open to wearing a helmet in the future.

To target those cyclists and others, the Ministry developed a communications strategy to strengthen voluntary helmet use among cyclists centred on those more implicit, emotional factors that do not only take rational cost-benefit calculations into account.
The research prompted the Ministry to adopt a holistic view of citizen decision-making when it comes to helmet use and ensure that emotional resistance does not override rational choice.

Key learning for implement phase, regardless of your policy area.

1. Adopt a holistic view of the environment the policy will be launched in and consider emotional as well as rational reactions.

2. Do not underestimate the strength of habits and social norms when a new policy needs to be adopted.

3. Design a sophisticated communications strategy not only relying on rational calculations when launching new initiatives based on evidence on what citizens will respond to.

The next example illustrates the challenge of ensuring stakeholder buy-in in the implementation phase. It is based on the experience of Ramboll in Germany.

How to ensure participation of intermediaries and target groups over a longer implementation period?

To improve teacher education during the course of study, the Federal Ministry of Education and Research (BMBF) launched a major funding and development program "Qualitätsoffensive Lehrerbildung". The aim of the program was to optimise education structures at universities, develop relevant support services for students and ensure comparability of studies and qualifications across the federal tates. The federal ministry commissioned Ramboll Management Consulting and the University of Linz to carry out a comprehensive evaluation in order to assess the impact of this large-scale program.

The program involved 72 universities across all federal states of Germany with 91 very heterogeneous projects. Therefore, a large and diverse group of stakeholders had to be involved and motivated to participate over a relatively long evaluation period. The evaluators relied on stakeholder cooperation, and the program also aimed to achieve cross-state cooperation in order to ensure impact and comparability.

To maintain the willingness of all stakeholders to participate in data collection over the entire period (2016 – 2024), Ramboll set up a large-scale evaluation programme including online surveys for all projects, monitoring systems, expert interviews, interviews with funded and non-funded universities and group discussions with representatives of the education and science ministries of the federal states, as well as observation at events and document analysis.

This wide variety of data sources ensured that all relevant data could be collected without putting too much burden on individual participants and that each participant was approached with a suitable research method. Very heterogeneous starting positions at the universities were considered through more extensive and less standardised monitoring instruments (e. g. more open response options).

The support of representatives of the federal states and a scientific advisory board underpinned the expertise and authority of the researchers. Moreover, the transparency of the evaluation results, for example through annual thematic interim reports and workshops to discuss interim results, played a key role in engaging stakeholders by increasing trust and already providing some benefits to universities

These measures ensured that the "Qualitätsoffensive Lehrerbildung" supported reforms and initiated new approaches to teacher education, as well as boosting research in teacher education. The projects and stakeholders were connected nationwide and supported in their work. The "Qualitätsoffensive Lehrerbildung" helped the universities to better manage the need for coordination and to become more resilient through networking.

> ## Key learning for implement phase, regardless of your policy area.
>
> 1. When targeting stakeholders, make sure your request is relevant and easy to fulfil for them
>
> 2. Highlight the benefits of the programme or research to stakehol ers.
>
> 3. Increase trust by being as transparent and approachable as possible.

The next example illustrates scaling challenges in the implementation phase after a successful pilot. The case study is based on the experience of testing simple behavioural interventions by the New South Wales (NSW) Behavioural Insights Unit to reduce the gender gap in recruitment for the NSW Government in Australia.

How to scale interventions that have been found to be effective during a pilot?

Since 2015, one of the Premier's priorities was to increase the proportion of women in senior leadership roles in the NSW Government sector from 33% to 50% by 2025. A key challenge was that women do not apply for senior leadership roles at the same rate that men do (this was uncovered in the Diagnose phase).

The Behavioural Insights team developed two simple, low-cost interventions designed to encourage women to reapply for senior roles after they narrowly missed out. The interventions were an email and a phone call with a recruitment manager that focused on how well the applicant had done and encouraged them to apply again. Both interventions leverage the behavioural tendency to judge an experience based on how we feel during its emotional peak and at its end. During the recruitment process, the emotional peak is often an interview, while the end for most candidates tends to be a generic email advising that their application was not successful. The intervention aimed to disrupt this final impression by prov ding candidates who were interviewed but not successful with an encouraging email or phone call.

The interventions were tested in a pilot among a group of public sector agencies in 2019/2020. A randomised-controlled trial design with 1614 candidates was implemented. These were randomly allocated to receive either a 'business as usual' recruitment process or an encouraging email or an encouraging email plus a phone call.

Two weeks after receiving the intervention (or not), candidates were sent a survey to record their overall satisfaction with the recruitment process, intention to reapply and likelihood to refer a colleague to the NSW Government. Moreover, it was monitored whether candidates reapplied for another senior role in the NSW Government within three and six months later.

The pilot results clearly showed that both interventions were successful. They had a significant impact not only on the intention but also on the behaviour to reapply for senior roles. Moreover, the impact was significantly larger for women than men, so that the gender gap was significantly re uced to 12% (email only) and to 4% (phone and email) compared to the control group with a gender gap of 45%.

It was calculated that if the email and phone call were scaled across the entire NSW Government, it could increase the proportion of women in senior roles by up to 0.5% each year and make a significant contribution to achieving equal gender representation in senior roles. Therefore, the Public Service Commission developed a toolkit to help recruitment managers implement the interventions in all clusters. Since then, gender diversity on NSW Government boards has further improved, although it has not yet been evaluated whether this progress is based on the roll-out of this treatment.

Percentages of female senior leaders in the NSW Public Sector
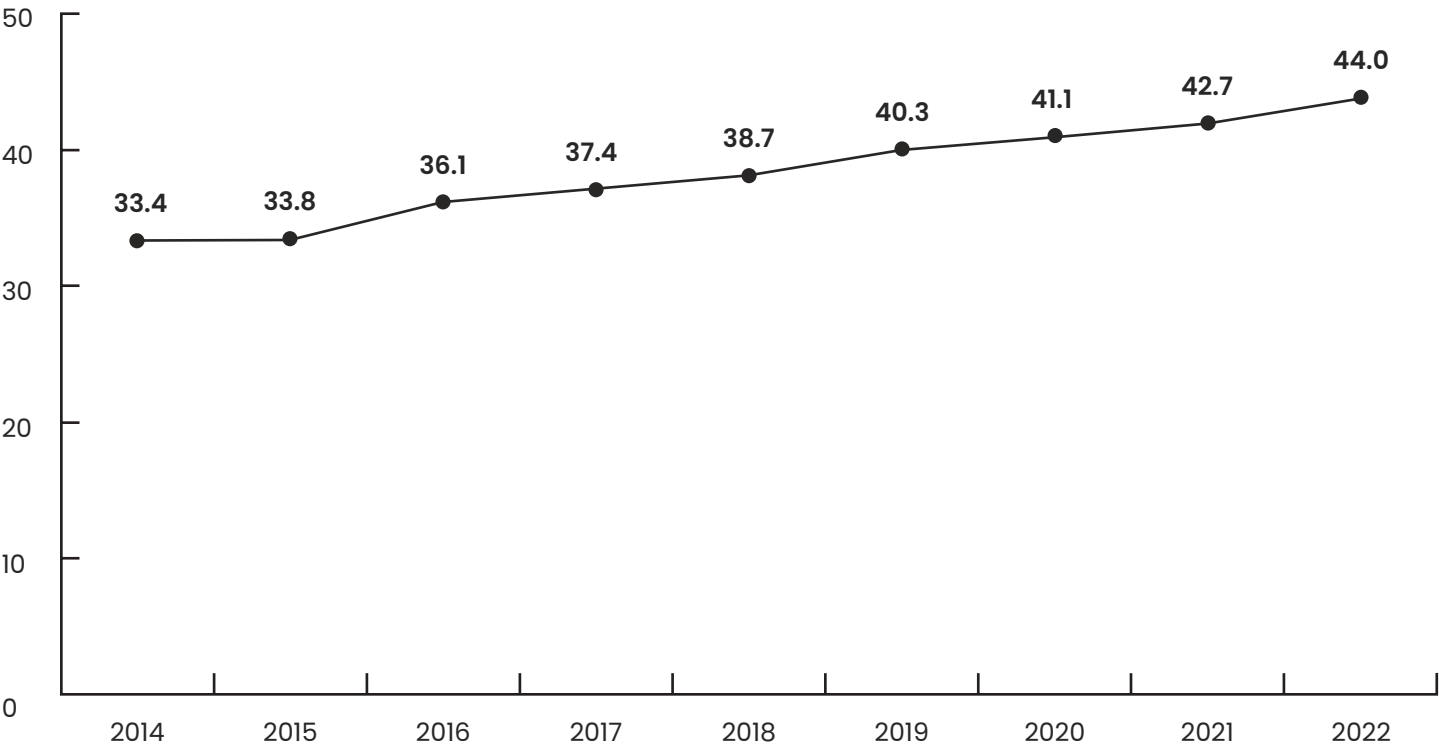
Source: NSW Public Sector Commission 2022, Workforce Profile Re ort 2022, NSW Government

Key learning for implement phase, regardless of your policy area.

1. Scaling interventions that have been found to be effective during a pilot is always difficult becaus it requires system change. It is critical to get buy-in for the pilot from those who will be involved in scaling. The costs of running pilots like this is low when administrative data are used for the evaluation, but agencies don't always have good data systems, and some technical expertise is required to conduct rigorous evaluations.

EVALUATE: Evaluating the output, outcome and impact of a solution

Without any monitoring and evaluation after the solution has been implemented, you do not know whether and how the solution has worked.

Collecting evidence during the implementation process will help as an early warning system if something goes wrong and needs to be corrected. Impact evaluations after implementation will give you the numbers that you may need to tell a success story, to understand whether it was value for money and to inform future policy decisions. The following research questions are typically asked at this stage:

Evaluate What worked and what didn't and why?	Has the selected option achieved the desired objectives? > Were the intended target groups reached to a sufficient extent? > Were there any unintended effects? > Was it value for money? How can you use monitoring & evaluation data to continually improve and engage and achieve sustainable change?

The first example below illustrates the challenge of designing   omprehensive and useful research programmes in the evaluation phase. It is based on the experience of Verian in Germany.

Has the selected option achieved the desired objectives?

The goal of the German education program "Bildungsprämie" was to increase participation in vocational training, especially for those who were hindered by financial barriers.   y offering financial support, the aim was to make training affordable for those on low incomes, addressing both cost and other barriers, such as limited time and skills.

To assess whether these effects were achieved, a research consortium consisting of Verian, Institut für angewandte Wirtschaftsforschung e.V and the Forschungsinstitut Betriebliche Bildung (f-bb) conducted a large-scale evaluation study with a holistic view of research aspects and questions.

While the evaluation focused on the standard question on effectiveness and goal achievement – "Did the programme increase participation in vocational training?" – a more comprehensive view was taken and also considered:

- Analysis and assessment of economic viability
- Analysis and assessment of educational success
- Identification of improvement potentials or fault controls in t e process
- Development of actions and control measures for the federal program

The research also aimed to evaluate any long-term impact and insights for future training promotion programmes. In doing so, the research took a broader view and not only measured the continuing education and professional development behaviours and attitudes of participants, but also the broader relevance of subsidised training and its effect on social compensation, mobilisation and the reduction of unequal education participation.

Consequently, it considered the embedding of the programme in the whole social context and further relevant developments.

This broad focus helped to identify a range of possibilities for improvement and for a sustained demand-oriented promotion of professional development, but also revealed limitations of the programme's effectiveness. It showed that participants were more likely to attend higher quality training or additional training, and it especially seemed to encourage lower-income individuals. Also, the participants reported high satisfaction with the "Bildungsprämie".

However, further analysis showed that possible effects of the education premium seemed to be due to selection bias, with only more motivated participants taking part in the first place. The impact on economic and career advancement also was modest. Considering economic viability, it turned out that the programme made individual funding very expensive.

A well-designed evaluation study with diverse and extensive research questions therefore managed to produce a well-balanced and nuanced assessment of its overall effects.

Key learning for evaluate phase, regardless of your policy area.

1. Make sure you capture all relevant aspects of a programme and its potential impacts when assessing whether it has been effective.

2. When assessing long-term effects, consider the social context and the broader environment as well.

The next example shows how monitoring systems can be useful in the evaluate phase to keep track of trends and new developments. It is based on the experience of Verian in the Netherlands.

How can you use monitoring & evaluation data to continually improve and engage and achieve sustainable change?

The Dutch National Sports Federation (NOC*NSF) aims to signific ntly increase sports participation in the Netherlands. To make sure all programmes are based on evidence and any new strategies are evaluated immediately, it commissioned Verian to conduct a monthly monitoring survey targeting a representative sample of Dutch citizens.

Each month, the survey measures sports participation and levels of physical activity, as well as the most commonly practised sports and the demographic make-up of Dutch sports participants.It relies on a standardised set of questions based on a validated methodology to assess sports participation and physical activity.

It can therefore produce a key indicator for sports participation that can be compared over time and across different contexts. As the data is collected monthly, a granular picture of trends and patterns emerges that allows for precise policy adaption.

Moreover, the large sample sizes amassed over this period allow for detailed statistical analysis and socio-demographic comparisons.

The monitoring survey also allows for ad hoc questions and special focus areas to obtain targeted data on emerging trends, seasonal activities or other particular areas of interest. The monitor therefore combines a standardised base module producing reliable and consistent outputs with the opportunity to flexibly adapt strategies and react to emerging opportunities in sports.

A fixed data pipeline feeds data into easy-to-access dashboards just days after fieldwork has ended and makes any insights available to use immediately. Being very fle ible, the dashboard incorporates feedback from previous months' data to make sure it is always up to date and produces timely and correct results. In general, the monitor is the basis for policy documents from NOC*NSF and the formulation of long-term strategies targeting the barriers identified in the survey for a diverse r nge of subsets of Dutch citizens.

Key learning for evaluate phase, regardless of your policy area.

1. When setting up monitoring systems, allow for both standardised and comparable measures as well as flexibility to react to trends and new focus areas

2. A design with several points of data collection helps understand trends and provides large samples to perform more detailed analysis.

Bibliography

- Agence de la Transition Écologique (ADEME). (2023). Scénarios avec et sans consigne pour recyclage des emballages de boisson (Scenarios with and without deposit for beverage packaging recycling). Evaluation report, published by ADEME.

- Hadorn, S., Sager, F., Mavrot, C. et al. (2022). Evidence-Based Policymaking in Times of Acute Crisis: Comparing the Use of Scientific Knowledge in Germany, Switzerland, and It ly. Polit Vierteljahresschr, 63, 359–382.

- Houses of Parliament. (2017). Parliamentary Research handbook.

- Ivert, A.K., Mellgren, C., Nilsson, J. (2020). Processutvärdering av Sluta Skjut (Process evaluation of Cease Fire). Evaluation report published by Malmö University.

- Kennedy, D. M. (2001). Reducing Gun Violence – The Boston Gun Project's Operation Ceasefire. Researc Report published by U.S. Department of Justice.

- Morestin, F. (2023). Consigne des bouteilles en plastique: le gouvernement recule devant une "fausse bonne idée" (Plastic bottle deposit: the government backs away from a false good idea). Article published by Novethic on 30th September 2023.

- Nationaler Normenkontrollrat - NKR. (2019): Erst der Inhalt, dann die Paragraphen. Gesetze wirksam und praxistauglich gestalten.

- Nesta. (2022). Engaging with evidence toolkit – A practical resource to strengthen capabilities for evidence use and expert engagement.

- OECD. (2021). Regulatory Policy Outlook 2021.

- Ofcom. (2023). Behavioural insights for online safety: understanding the impact of video sharing platform (VSP) design on user behaviour. Economic discussion paper series, issue number 4.

- O'Donoghue, K., Moore, A. (2022). Reducing Sexual Violence – research informing the development of a national campaign. Commissioned by the Australian Government, Department of Social Services.

- Riphahn, R.T., Schnitzer, M. (2022). Die Bedeutung wissenschaftlicher Evidenz für die Arbeit des Bundestags. In Vierteljahrshefte, 91(4), 271–272.

- Sager, F., Hadorn, S., Balthasar, A., Mavrot, C. (2021). Politikevaluation – Eine Einführung. Springer VS.

- Seidel, K., Verbeek, H., Fessel, S., Meer, F. (2021). Nutzen von wissenschaftlicher Evidenz – Erwartungen an wissenschaftliche Expertise. Bericht über eine Bundestagsbefragung im Rahmen der Leopoldina-Evidenzinitiative unter Leitung von Regina R. Riphahn und Monika Schnitzer. Diskussion Nr. 27, Halle (Saale).

- Social Policy Evaluation and Research Unit (Superu). (2018). Making sense of evidence: A guide to using evidence in policy. Wellington, New Zealand.

- World Health Organisation. (2021). Evidence, policy, impact: WHO guide for evidence-informed decision-making.