d·i·e

Deutsches Institut für Entwicklungspolitik

German Development Institute

Discussion Paper 8/2018

Evidence-Oriented Approaches in Development Cooperation

Experiences, Potential and Key Issues

Paul Marschall

Evidence-oriented approaches in development cooperation

Experiences, potential and key issues

Paul Marschall

Bonn 2018

Discussion Paper / Deutsches Institut für Entwicklungspolitik
ISSN 1860-0441

Die deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.d-nb.de.

ISBN 978-3-96021-064-1 (printed edition)

DOI:10.23661/dp8.2018

Printed on eco-friendly, certified paper

Dr Paul Marschall is a researcher in the research programme "Inter- and Transnational Cooperation with the Global South" of the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).

Email: paul.marschall@die-gdi.de

Tulpenfeld 6, 53113 Bonn
☎ +49 (0)228 94927-0
🖷 +49 (0)228 94927-130
Email: die@die-gdi.de
www.die-gdi.de

MEMBER OF JRF Johannes-Rau-Forschungsgemeinschaft

Contents

Figures

Tables

Acknowledgements

I would like to thank Stephan Klingebiel, Sarah Holzapfel, Sven Grimm, Anna Schwachula, Julian Bergmann and Heiner Janus for their helpful and valuable comments.

Bonn, April 2018 Paul Marschall

Abbreviations

3ie	International Initiative for Impact Evaluation
BCURE	Building Capacity to Use Research Evidence
BMZ	German Federal Ministry for Economic Co-operation and Development / Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung
CBA	cost-benefit analysis
CC	Copenhagen Consensus
CEA	cost-effectiveness analysis
DAC	Development Assistance Committee
DANIDA	Danish International Development Agency (brand used by the Ministry of Foreign Affairs of Denmark when it provides humanitarian aid and development assistance)
DC	development cooperation
DEval	German Institute for Development Evaluation
DFID	Department for International Development
EBM	evidence-based medicine
EBPM	evidence-based policy-making
GIZ	Deutsche Gesellschaft für internationale Zusammenarbeit
ICAI	Independent Commission for Aid Impact
ISS	immunisation services support
KfW	KfW Development Bank / Kreditanstalt für Wiederaufbau
LIC	low-income country
MDG	Millennium Development Goal
NGO	non-governmental organisation
ODA	official development assistance
OECD	Organisation for Economic Co-operation and Development
RBA	results-based aid
RBF	results-based financing
RBF4MNH	results-based financing for maternal and newborn health
RBM	results-based management
RCT	randomised controlled trial
SDG	Sustainable Development Goal
USAID	United States Agency for International Development

Executive summary

Background

The use of more evidence as an instrument for achieving higher impact in development cooperation (DC) is a major topic in current discussions on how to improve the impact of aid. Based on a broad understanding of evidence, this discussion paper is a contribution to answering three questions. First, how is evidence currently provided in DC? Second, what are ways of using evidence in this regard? Third, what is the potential of considering evidence in policy-making in the future?

Evidence

Evidence is a term with several meanings and connotations. In this paper, the definition takes a broader view as the ability to draw conclusions based on empirical available data, information and/or knowledge that an activity works. On an operational level, evidence corresponds to the concept of effectiveness, the relationship between planned and achieved results. Ideally, effectiveness is about "doing the right things". The conceptual framework for this analysis consists of three dimensions: 1) the comprehensiveness of different types and sources of evidence used (scope); 2) the strength and quality of evidence used (strength and quality); 3) the extent of using evidence in policy-making (evidence-orientation).

Evidence is based on different sources: (i) academic evidence – which is broader than scientific evidence, which only applies to the natural and social sciences – is ideally objective, replicable and generalisable and is provided as academic output; (ii) ministries such as the UK Department for International Development (DFID), implementing organisations such as Deutsche Gesellschaft für internationale Zusammenarbeit (GIZ), the administration, but also non-governmental organisations (NGOs) – in their function as DC implementing organisations – provide practical-technical evidence; (iii) social evidence refers to data and information directly and mostly contributed in an aggregated way by members of civil society by using, for example, social media, or through population movements. In general, scope relates to the number of evidence types considered. There are, for example, different hierarchies used by academics for ranking the strength and quality of provided evidence. It is useful to extend such basic schemes to also cover evidence from other sources. Finally, it is assumed that policy-making rests on opinion and/or on evidence. The extent to which policy-makers consider evidence (evidence-orientation) is defined by the levels of opinion-based, evidence-supported and evidence-based policy-making.

Providing evidence in development cooperation

A broad spectrum of sources and instruments can be used for generating and disseminating available evidence. Unstructured data forms the foundation of evidence. However, the value of data is limited before being processed and analysed. Special formats such as academic journals, reports, other papers and newsletters are used for providing and channelling academic, practical-technical and social evidence. There are platforms for knowledge-exchange. Frequently, the creation of data and information is a by-product of DC. Governmental and international organisations – and increasingly, non-governmental providers of DC – collect monitoring data. At the partner-country level, statistical data are collected.

Household surveys are an important source of evidence. In the last decades, monitoring has also become an important instrument for presenting the performance of projects and programmes. In addition, performance indicators enable measuring the progress in achieving development goals, currently in terms of the Sustainable Development Goals (SDGs).

In order to improve the results-orientation of projects and programmes, existing management approaches were adjusted recently. Results-based management (RBM) is a concept with a strong focus on performance and achievement of outputs, outcomes and impacts. Operational evidence is an important key source for implementing learning processes. A few years ago, results-based approaches were introduced. They include a variety of approaches aiming to shift the paradigm of aid from a focus on inputs and activities to a focus on results by reshaping the relationship between development partners/ donors and partner countries. Therefore, quantifiable and measurable results must be identified that are linked as closely as possible to the effects of DC. For example, the immunisation services support (ISS) provided by GAVI, The Vaccine Alliance, helps countries to expand routine immunisation coverage based on performance payments calculated from immunisation data provided from countries' administrative reporting systems.

Recently, different evaluation techniques beyond project or programme performance evaluation have attracted attention. Impact evaluation aims to determine mostly the longer-term results generated through policy decisions, often through interventions, projects or programmes. Impacts may be positive or negative, intended or unintended, direct or indirect. In particular, randomised controlled trials (RCTs) – a methodological approach often used in evidence-based medicine (EBM) and increasingly in DC – is often regarded as the "gold standard", enabling high-quality evidence. However, RCTs also have their limitations, in particular in social settings, where every trial group is different. Once trials were conducted, review groups became biased. Thus, these shortcomings must be considered in systematic reviews, which are powerful instruments for synthesising available evidence.

Use of evidence in development cooperation

The use of evidence for analytical and operational purposes depends critically on its available quantity and quality. Moreover, the time frame and the irretrievability of sources matter. Until recently, evidence about what works, both in development and DC, was frequently not collected systematically. Only particular evidence related to research outcomes on special geographical settings and research areas was available. Results management and results-oriented approaches were just starting, and the available evidence about project and programme performance was poor. Overall, the quality of evidence was low, and the corresponding value for decision-making limited. In the last few years, the quantitative and qualitative potential of evidence in DC as input in policy- and decision-making has improved a lot. The number of evaluations and their quality have increased. International networks and organisations now provide capacity-building for disseminating more and better evidence. There are institutes and persons who aim at awareness-raising for using more and better evidence at the level of decision-making and administrative bodies. However, there is still potential for improvement, for example in terms of including more accompanying research.

Evidence matters in policy-making. It is used twofold: 1) symbolically, for increasing the credibility of the decision-makers and their decisions, and 2) instrumentally, to adjust knowledge

and improve decision-making. Because of a strong push from national and international initiatives, the awareness of the value of evidence for DC has risen in a striking way, but the transmission of knowledge and the adoption of evidence in different settings is rather mixed. The demand for – and use of – evidence differs a lot. Pioneering countries such as Australia, Canada, New Zealand and the United Kingdom have pushed the issue on the agenda and are working on evidence-oriented capacity-building also in partner countries. A major topic is impact evaluation, which provides both information for improving running activities and designing new projects. The gains from providing evidence through results management have, by far, not been exploited yet. This is also caused by the complexity of results frameworks and measurement problems. Nevertheless, there are also other barriers preventing evidence from being used. Policy-makers are often political actors. Thus, values, political beliefs and the interaction with the political system are relevant. In practice, evidence is only one among a number of factors at play in policy processes. Sometimes, available evidence is not considered in decision-making because evidence is not available when needed. Among other reasons, whether or not policy-makers consider evidence critically depends on the perception of whether the provided knowledge is perceived as a solution to an existing problem or not.

Case study: the Copenhagen Consensus

There are still enormous challenges in translating academic evidence into practical use due to less appropriate transmission formats. Currently, social media and marketing campaigns, as used by the Copenhagen Consensus (CC) Centre, are important instruments for attracting attention. The latter is an international Copenhagen-based think tank established in 2002. Since 2004, several "projects" have been conducted that have focussed mostly on development problems at the regional, national and global levels. Optimal solutions for the problems under review have been identified by using the routine economic concept of cost-benefit analysis (CBA), a valuation technique that is used for ex post evaluations. In CBA, the net benefit of investments is calculated by comparing the expected returns from alternative policies (= value for money). The ultimate logic behind this is to choose the more effective ones. An analysis of the CC approach is revealing, as it can be understood as a prime example of academic evidence and the corresponding barriers for using academic insights in policy-making. The CC concept is highly controversial, being a project dominated by economists. Many academics acknowledge the value of the background reports, which are used as an important input for a CBA, but they disagree with some of the models and assumptions used in a CBA. For example, many academics working in the field of environmental sciences rejected the CC results due to the way in which climate threats were methodologically implemented. The CC results were also criticised for breaking down independent silos of funding and priorities by offering a simultaneous choice between a variety of problems and solutions.

Potential of evidence in development cooperation

Pathways to success are based on accompanying measures, including ongoing policy advice. It is important to assess available evidence critically because it is not always consistent and implementable. Because of the continuous contact between suppliers and demanders of

evidence, appropriate evidence is identifiable and provided. For understanding why perceived evidence is not used in policy-making and implementation, the political economy must be considered. In spite of international agreements aiming at more aid effectiveness and lip service about the "importance of more evidence-based policy-making", political actors often are not interested in taking the whole menu of provided evidence into account. Ranking schemes – including the value-for-money of different interventions, as provided by the Copenhagen Consensus Centre – are welcomed by policy-makers. Such menus are often used for "cherry picking". Existing uncertainty about the future and serious doubts towards underlying calculations provided by scientists and think tanks can also easily be used as excuses for not making decisions based on evidence.

The ongoing discussion about including more evidence in DC is mainly focussed on academic and practical-technical evidence. The direct inclusion of social evidence and civil society's experiences, in particular, is dependent on the setting. For example, there is evidence that it is sometimes used in the health sector because people are understood to be important stakeholders. At the high level, it is considered by incorporating (representatives of) civil society in policy-related decision-making. At the operational level, there is more space for taking social evidence into account.

The practice of DC shows that the "call for more academic evidence" is often too narrow. Strong and intensive collaboration between the operating units of implementing organisations and academics and an open dialogue can provide a win-win situation for all partners and a guarantee for valid evidence, which is an important source for learning at the operational level. Over time, the strength and quality of available evidence has improved in a striking way. For example, there are now many evaluations available that rest on rigorous methods.

In comparison to the past, the level of usage in DC has increased substantially. But the call for evidence by political decision-makers is sometimes motivated by the decision to justify decisions already taken. In more advanced settings, the understanding and perception of the value of evidence is probably stronger. In those settings, the available quality of evidence might be better, and people – in particular those with an academic background – grow up in a more evidence-oriented culture. However, this does not mean simultaneously that political processes will automatically absorb evidence. It must be provided in an adequate way, for example through co-production with the relevant stakeholders. Their numbers and the homogeneity of their (special) interests are also important issues. In the case of for-profit companies with a predominant focus on maximising profits, the use of evidence about that issue will probably be greater than in policy-making in a multi-dimensional political space. Political priorities also matter. Thus, the basic allocation of funds from DC is different from value-for-money considerations, as suggest by the Copenhagen Consensus. To sum up, current policy-making can be understood as evidence-supported. Increasingly, NGOs also perceive the importance of using available (selective) evidence in their demands.

Similar to areas of activity such as health and education, the importance of evidence in DC will increase in the future. In order to improve the concrete level of relevance of evidence in policy-making, the existing creation and provision of evidence should be strengthened. Types and forms of evidence transmission for practical use have to be improved, and policy-makers must acquire a better understanding of the usefulness of evidence, at least for reasons of accountability and the better performance of activities targeted at their political goals. For

this, an evidence-oriented culture based on transformative research integrating all societal actors in the process of research by co-design and co-production of knowledge is extremely helpful.

Conclusions and recommendations

Based on the results of the above theoretical analysis and the available experiences, some important recommendations for German DC can be given.

To increase the impact of German DC, a stronger focus on evidence is needed. In general, it is helpful to establish a strong evidence- and results-based culture in all parts of the German DC system. There is also a need for more systematic learning. This can be supported by improving institutional structures.

The objective should be to incorporate evidence into decisions in a timely and efficient way. Based on the particular issues of problem-setting, including the timescale, the types and pieces of required evidence should be identified and collected.

Striving to identify "what works" is central to the mission of German DC and to those who are supported. The type of evidence used, as well its strengths, should be specified when making or proposing decisions. Reflecting upon and defining the proper criteria for assessing the evidence's strengths, such as validity, timelines and reliability, can be helpful.

For supporting evidence-oriented policy-making, specific studies and evaluations can be beneficial. Available evidence must be the backbone of policy-making. In addition, results management is important for accountability and improving decision-making and needs to be strengthened.

Decision-makers in partner countries must be supported in using evidence in policy-making. To improve the impact in neglected policy areas, existing evidence gaps must be identified and addressed. To increase the impact of actions, providers must improve their networking and the pooling of available results.

1 Introduction

The recently published Development Cooperation report of the Organisation for Economic Co-operation and Development (OECD, 2017) has a particular focus on "data for development". The authors argue that there is a need for a better evidence base for making informed choices about priorities and strategies to achieve the Sustainable Development Goals (SDGs).

"Evidence" is a dazzling term. In science, the word has been used for hundreds of years. The Latin root (*evidere*) reminds us of bringing to the fore what is there to be seen. The term also has become apparent in practice and "real life" in common language. To some extent, "evidence" is now a buzzword with different connotations. What constitutes evidence is often not clear-cut. Evidence is used as an umbrella term, either interchangeably with "empirical information" as such, or in a more narrow sense, as in the "ability to establish or support conclusions" because there is an available body of facts or information that indicates whether a belief, proposition or theory is obviously true or valid.

In a more concrete sense, it is often unclear for the audience and the reader about what a presenter exactly means by the expression "evidence". Sometimes, the meaning only becomes clear when the concrete context is considered.

The world of evidence is also paradoxical. Evidence rests on unstructured data, which need to be processed and analysed. Even the best statistical offices are not able to cover all SDG targets adequately. Data gaps are still enormous in some developing countries. However, "big data" is also an important issue. In order to produce evidence, available (meta) data must be analysable. Currently, there is an intensive discussion about data mining, including data privacy.

Looking into the literature, there seems to be a global push for evidence-based policy-making (EBPM) (OECD, 2017), though there are divergent views on what evidence-informed policy actually is, and how it relates to policy influencing agendas (Hewman, Capillo, Famurewa, Nath, & Siyanbola, 2013). A variety of evidence-oriented approaches were introduced in many policy areas and disciplines, such as psychology, criminal justice, nursing, education, social care, transport, and library and information science (see e.g. Davies, Nutley, & Smith, 2000; Young, 2013). However, those new discussions were decoupled to a broad extent from the original philosophical debate (Achinstein, 2001), the basic root of evidence. Important philosophers such as Franz Brentano and Edmund Husserl argued that evidence is not relocatable. Because of its direct connection to truth, evidence always has an absolute meaning. In contrast, understanding the different levels of hierarchy and the corresponding consequences is central to applications in evidence-based medicine (EBM) or for the use of evidence in development cooperation (DC).

The discussion about "more evidence" also has a strong link to observable changes towards informed societies, the omnipresence of data and information, and some kinds of facts. The call for more evidence-oriented decisions also arises from the shortcomings of existing concepts of decision-making and the disappointing results. There are hopes that decisions which are more evidence-informed will improve outputs, outcomes and impacts and will also save the scarce amount of money available. There are some indicators that the provision of evidence by academics is continuing to grow. Among others, the number of academic

journals providing evidence continues to grow, and many studies, analyses and evaluations are now widely available on the internet.

Much of the evidence-providing literature has a clear focus on the academic community, but its practical use is limited due to the restrictive way it is presented. Many reports addressed directly to policy-makers actually do not reach the intended level. For example, the World Bank invests about one-quarter of its budget for country services into knowledge products. Recently, researchers found that almost one-third of the World Bank's archived policy reports – documenting the impacts of its numerous projects, from dam construction to micro-crediting – have never been downloaded (Doemeland & Trevino, 2014). The corresponding problem is not limited to the World Bank.

Evidence-based policy-making and practice is not a new topic in DC. This concept was already included in the UN Millennium Development Goals (MDGs) guide: "Evidence-based policy-making refers to a policy process that helps planners make better-informed decisions by putting the best available evidence at the centre of the policy process" (United Nations Development Programme, 2007). However, just recently, the call for more evidence has triggered a more intensive discussion in DC.

It is the main objective of this paper to analyse the provision and use of evidence in this policy area. Based on a broad understanding of evidence, this discussion paper is a contribution to answering three questions. First, how is evidence currently provided in DC? Second, what are ways of using evidence in this regard? Third, what is the potential of considering evidence in policy-making in the near future? This refers to how – and in what context – it can be used.

Therefore, in Section 2, a comprehensive framework for studying different categories of evidence is developed, including types and forms of evidence, quality and strength. Furthermore, the question of assessing the use of evidence is studied. In Section 3 an overview of different forms of evidence in development cooperation is given. Section 4 presents some information about the current use of evidence in DC. As a case study for providing evidence for policy-making, the approach of the Danish think tank Copenhagen Consensus Centre is presented and discussed in Section 5. Subsequently, the impact of the existing evidence in development cooperation is analysed. Finally, Section 7 draws some conclusions about perspectives on evidence in this policy field.

2 Evidence

2.1 Background and context

The roots of the conceptual understanding of evidence can be traced back to early writings in the philosophy of science and epistemology, emphasising a precursory stage of verity. In a very broad sense, the term expresses that there is an available body of facts or information that indicates whether a belief, proposition or theory is obviously true or valid. This corresponds to effectiveness, which is the relationship between actual and planned results.

There is a call for more evidence in many action fields and policy areas in a dynamic and changing social, political, environmental and technological context. Appendix 1 presents some examples. For many years, there have been initiatives for fostering closer and more

effective links between research and policy (Oliver, Lorenc, & Innvær, 2014). The field of medicine and public health became the most promising discipline. Based on initial thoughts in the 19th century (Hjørland, 2011) and seminal works on effectiveness and efficiency in 1972, the concept of "evidence-based medicine" was introduced in 1992 (Evidence-Based Medicine Working Group, 1992). This became something like an offshoot in thinking about the value of evidence in disciplines such as sociology and political science. However, there are important differences between implementing EBM and providing evidence in another context, for example in policy-related issues on larger-scale decisions (Pawson, 2006).

In general, there are two basic strands of evidence-oriented discussions. One of them is practice-oriented.

- *Strand A* focusses directly on the micro-level of *practice*. Programmes or practices are evidence-based if they effectively integrate the best research evidence with some kind of practical expertise and – where applicable – cultural competence and the values of the persons receiving the services. Example: evidence-based medicine.

- *Strand B* aims to transfer evidence from different sources to the *policy* level. Frequently, this discussion is limited to academic evidence. However, some other sources and types of evidence matter. Example: evidence-based public health policy.

A number of factors contributed to the rise in the role of evidence in policy and practice: among others, the growth of an increasingly well-educated and well-informed public, the explosion in the availability of data of all types, the growth in the size and capabilities of the research community, and an increasing emphasis on productivity and international competitiveness (Nutley & Webb, 2000). In DC, the growing awareness about "aid that works" was pushed by the MDGs, the Monterrey Consensus and the Paris Declaration on Aid Effectiveness. Many large non-governmental organisations (NGOs) also recognised the need for broader interventions to achieve development goals.

There are many different perceptions of what evidence is and how it should be used in policy-making and in practice. It is generally accepted that there is no single type of evidence that can provide all the answers. Different types or dimensions can and should be distinguished. For example, Tytler (2001) makes a distinction between three dimensions of evidence, as follows.

Formal academic evidence rests on academic results. Frequently, theories and models are starting points for research questions. Sometimes, hypotheses are tested based on a dataset. There is much research on identifying causal relationships between variables for drawing conclusions. EBM is an example. It has a focus on concrete diseases and rests on many academic study results in this regard. Based on the information that a concrete (drug) intervention might be beneficial for patients with certain characteristics, recommendations were written (guidelines). Evidence is the proof showing under what circumstances a therapy might work. Unlike formal evidence, *informal evidence* refers to common sense or personal experience, which lies outside the restricted scope of academic evidence. Anecdotal evidence is a similar term. A third type of evidence contains *wider issues* that are influenced by evidence, such as *environmental or legal concerns*. In jurisdiction, evidence is the means – sanctioned by rules – of ascertaining the truth respecting a matter of fact in a judicial proceeding.

In the literature, some other evidence classifications are suggested. The main problem of those typologies is that the types of evidence included are often not clearly differentiated. For example, Buetow and Kenealy (2000) draw a distinction between scientific evidence, theoretical evidence, practical evidence, expert evidence, judicial evidence and ethics-based evidence. Those categories overlap to some extent. It can also be questioned whether a more detailed subdivision is really helpful. For example, in practical and expert evidence, some informal evidence can also be included. In addition, offering a situational understanding of practitioners or experts as whole persons in their own environments and the nature of their social interactions can also be helpful. An important issue provided by Buetow and Kenealy (2000) is their argument that many questions have different facets. Even rather technical decisions, especially when social phenomena are covered, cannot be placed in a moral or social vacuum. This is an important issue in DC. Interventions may not only be guided by thinking that is grounded in European culture and values. They must also fit the culture of the intended beneficiaries. For sound decision-making, it is important to identify which types of evidence are relevant. Regression analysis is an important method for identifying evidence. However, one of the weaknesses of regression analysis is that corresponding evidence is based on the average of the total sample. Actions based on average-based evidence can be for the benefit of the majority of a population, but it is possible that vulnerable groups are excluded. This is exactly the point raised by the SDGs when they refer to leaving no one behind.

2.2 Conceptual framework

Unmanipulated data of a certain quality level are used as facts. They represent a description of the real world. Data are building blocks used to form knowledge. In addition, they can be captured in information so that other people can access them at different times. Information is not equal to knowledge because available information must be processed and absorbed. It has been argued that knowledge is the product of knowing and can only reside in the personal domain. Only information, theory, experience and research can be transferred from person to person (Scott-Findlay & Pollock, 2004).

Knowledge is based on theoretical and/or empirical insights. However, knowledge must not be considered to always be true. Our understanding can rest on wrong or biased data. Difficulties in interpreting data can lead to biased knowledge. Learning is a process by which the accumulated knowledge capital is extended. It is possible to acquire know-how either empirically through (own) experience and/or observations, or through theoretical insights.

Evidence is not synonymous with know-how, even though – according to the evidence type – the contours blur. In this concept, it is assumed that evidence is a subset of the latter, in which four distinct criteria are met:

1. Evidence only includes empirical-based know-how.
2. In the case of evidence, the data and information used are collected in a more systematic way.
3. There is a kind of judging of the trustworthiness of the given information.
4. Consequently, the claim that “there is evidence” rests on a minimum level of quality.

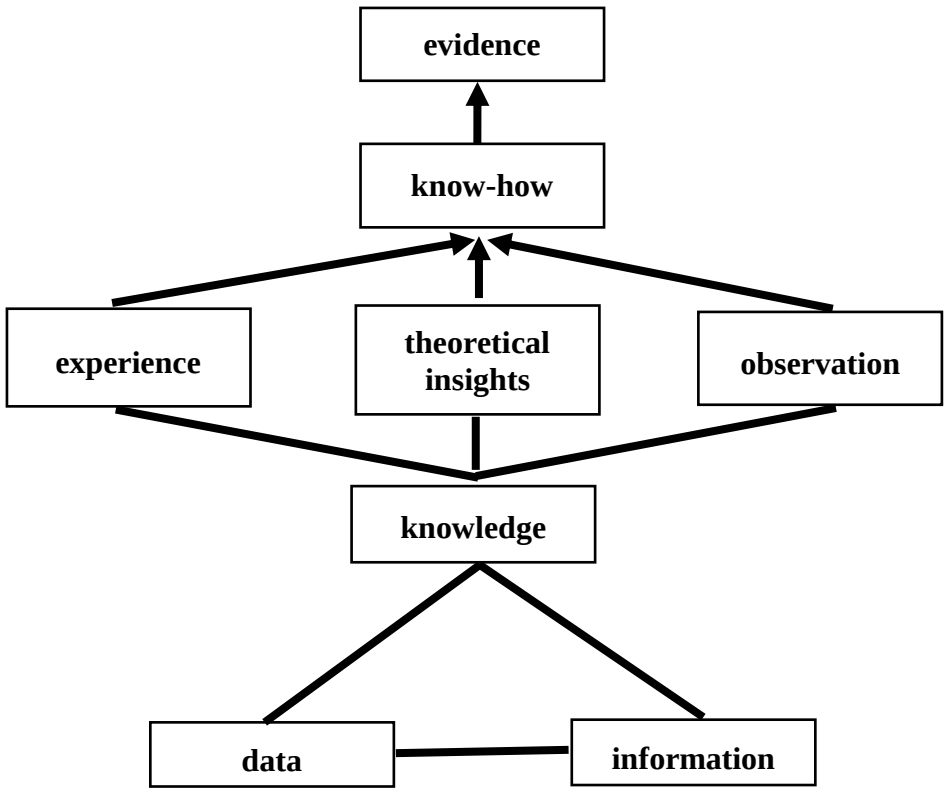
This refers to the used data and information, or methods used for creating know-how. "Empirical evidence" is an excellent example of that. Frequently, statistical methods are used to study whether a given statement is supported by data or experiences. For example, a given dataset is used for hypothesis-testing. Based on a chosen significance level, it is calculable if there is statistical significance at a certain level. Based on this example, it also becomes clear that the relevance of a statement is limited to the given dataset. In the case of a small dataset, it is not possible to conclude immediately whether or not the relationship is also valid for a more comprehensive dataset. Poor data quality or some outliers can create a high level of bias. This example also demonstrates that know-how believed to be evidence must be used and interpreted very carefully. Figure 1 presents the building blocks of evidence.

Decision-makers use a certain scope of evidence. The used information base will not consist of a homogenous set of information, but rather of several different pieces that are not always coherent and consistent, combined and weighted according to the decision-makers' preferences or values. In general, there are incentives for a stakeholder to cite information from the past as being "evidence" because, under a common understanding, "evidence" has the connotation of "some kind of proof". There is also the danger that evidence is misused, that is, suppressed or used to serve a political objective (Witter, Kardan, Scott, Moore, & Shaxson, 2017). The individual stakeholder determines whether or not there is a need for "proof of evidence". In this regard, it must be distinguished whether evidence is objective, or rather subjective.

The process and procedures of evidence-seeking can vary widely according to the relevant concept of evidence that is being used. For example, the discussion about EBPM was inspired by EBM, but the focus is qualitatively different. EBPM targets entire populations, whereas EBM is patient-oriented, and the decisions associated with EBPM are generally subject to greater public scrutiny (Dobrow, Goel, & Upshur, 2004).

The underlying logic of the EBPM approach is attractive to many theorists and practitioners, who agree with the premise that the better decisions are those that incorporate the best available information (Howlett & Craft, 2013). However, decisions are not taken about the past, but in terms of the future. Thus, there is not only uncertainty about the "value of evidence" itself, but also uncertainty relating to different states of the future. Evidence from the past must not be valid in the future. To translate knowledge, which is historical in nature, into the future, an adequate model is needed. In addition, we are living in a dynamic environment and in an era of rapid changes and increased uncertainty that is full of complexity and non-linearity. Evidence used for configuring strategies and interventions must adapt knowledge from the past in an adequate way. Thus, it is probably not adequate to update evidence in a linear way.

Figure 1: Conceptual roots of evidence



data	given facts; they are a description of the world; they represent material available for analysis and interpretation
evidence	there is an available body of facts or empirical information that indicates whether a belief, proposition or theory is obviously true or valid
experience	practical knowledge, skill or practice derived from participation in events or in a particular activity
information	data organised into meaningful unions and placed in context with relevance and meaning
know-how	knowledge of how to do something well
knowledge	a collection of all that is known
observation	an act of recognising and noting a fact or occurrence often involving measurement with instruments
theoretical insights	knowledge without empirical basis

Source: Author

For making decisions, policy-makers often combine many different pieces of information based on a diverse set of information sources of varying quality and weight them to generate action alternatives. Policy-makers are frequently not only interested in strong facts provided by researchers. Corresponding results are often interpreted against the background of their own (political) values and opinions, the key supporters, the interested and affected stakeholders, and the general public (Lavis, Ross, & Hurley, 2002). Policy-makers are an extremely heterogeneous group. They include, among others, government officials, legislators, civil servants, the judiciary and the media. In addition, these actors are engaged in a wide range of activities (Oliver, Lorenc, & Innvær, 2014). According to the understanding of political economy, decision-makers want to maximise the probability of re-election or of being elected in the next election. Therefore, they have an interest in using the kind of information that supports their beliefs. From the viewpoint of political decision-makers, "best evidence" does not necessarily mean "best quality of evidence", but rather "best evidence in line with own political beliefs". It is not important that the used evidence is valid and of high quality, but rather what voters anticipate. The political economy of natural disaster aid is a prominent example (Cohen & Werker, 2008). From the social science perspective, natural disasters consist of two different components: the perception of an environmental disturbance as such, and the corresponding assessment. In order to understand a disturbance as a "natural disaster", a critical threshold level must be exceeded. The media can construct the corresponding evidence indicating that policy-related actions are required (Miles & Morse, 2007). Based on that, political decision-makers receive information that, for example in the case of an earthquake in Haiti, voters support providing humanitarian aid and will reward that decision (Lundahl, 2013).

For politicians, that is, persons who are professionally involved in politics, especially as holders of an elected office, the public perception of own statements and actions taken is of high relevance, in particular shortly before elections. Available evidence can be useful if it confirms that certain policies are successful. Evidence can thus be used as an instrument for justifying own plans and activities. Thus, there is a danger in commissioning a research project to support a policy that has already been decided upon, because "evidence" as such also has a strong convocational appearance (Marmot, 2004). However, politicians also have a real interest in evidence being used as a feedback instrument for correcting policies that are not well-designed. Because evidence is mostly associated with uncertainty, developing further actions on top of it does not automatically guarantee success. Thus, politicians are free to use evidence in a strategic way, provided that evidence is also being questioned, as done by US President Donald Trump regarding the question of climate change. Evidence may not be confused with truth or verity. Because of existing shortcomings in providing evidence and uncertainty, basing decisions on it may be wrong. In addition, as Jasanoff (2006) argues, scientific thinking is always a social enterprise that is embedded in particular settings; it is context-specific, purposive and culturally embedded. However, the rejection of evidence should be well-founded. In addition, politicians can take advantage of the fact that informal evidence must not be based on rigorous methods.

Figure 2 presents a framework for analysing the provision and use of evidence. The cube of evidence consists of three dimensions:

(1) the scope of evidence

(2) the strength and quality of evidence

(3) the extent to which evidence is used in policy-making

All dimensions are explained in the following.

(1) The scope of evidence

The vertical axis of the evidence cube describes the scope of evidence used. Here, the variety of evidence types consists of three broad categories:

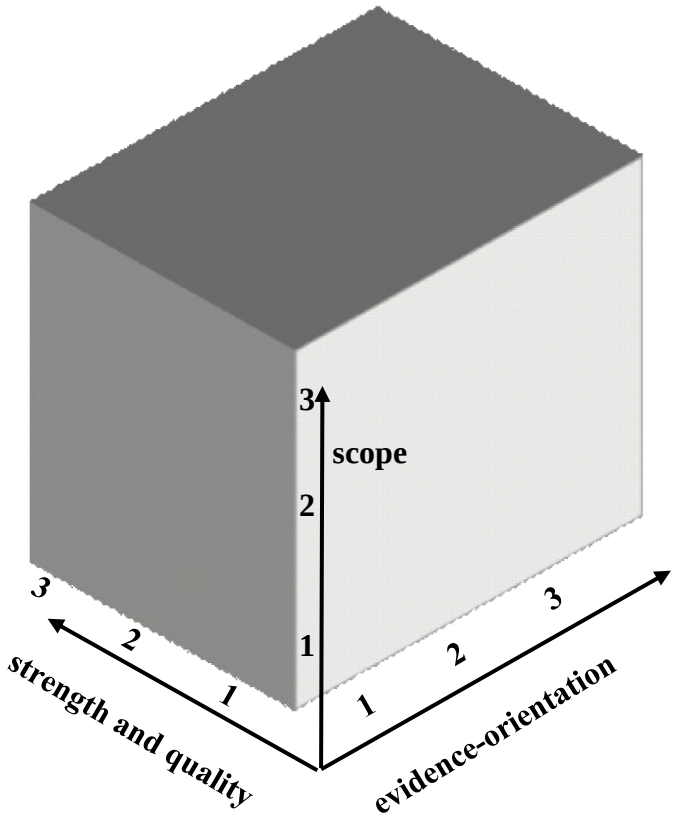
1. *Academic evidence* draws upon the empirical findings of all disciplines, including philosophical and ethical evidence, and all policy areas, such as education and health. In the *ideal* case it is:
   a. objective, meaning that it is observable by others, based on facts and free of bias or prejudice that might be caused due to personal opinions;
   b. replicable, in the sense that others can reproduce results by using the same methods used originally;
   c. either generalisable, meaning that it can be allied to other settings, and/or useful for internal learning processes, as in case studies with a narrow context-specific setting.

   This definition is not in conflict with the understanding that social processes colour the extent to which pieces of scientific knowledge are perceived as being certain (Jasanoff, 1987).

2. *Practical-technical evidence* includes know-how from administration and ministries – including the Department for International Development (DFID) and the Ministry of Foreign Affairs of Denmark (DANIDA) or DC implementation agencies such as the French Development Agency (AFD), Deutsche Gesellschaft für internationale Zusammenarbeit (GIZ) and Kreditanstalt für Wiederaufbau (KfW) – but also from NGOs that work in DC. Often, instruments used for providing this type of evidence rest on methods developed by academics.

3. *Social evidence* is based on the aggregated experiences and perceptions of individuals. People are, for example, the beneficiaries of projects or programmes. They communicate with each other about their ideas and reflections. Social evidence also includes contributions from civil society provided through instruments such as petitions and manifestos but also common sense or the personal experiences of civil society members, which lie outside the restricted scope of the academic and practical-technical spheres. Social evidence provides additional information about the credibility and suitability of other types of evidence. Whether, how and to what extent civic epistemology is considered depends on the style of public knowledge-making, which differs by national and cultural contexts (Jasanoff, 2005). Among other things, this type of evidence refers to information from civil society about perceptions of using public money via report cards (Open Budget Initiative), using mobile phone apps for reporting corruption, or information based on protests or public consultation sessions. Social networks have become important instruments for communicating social needs.

Figure 2: The nature of evidence



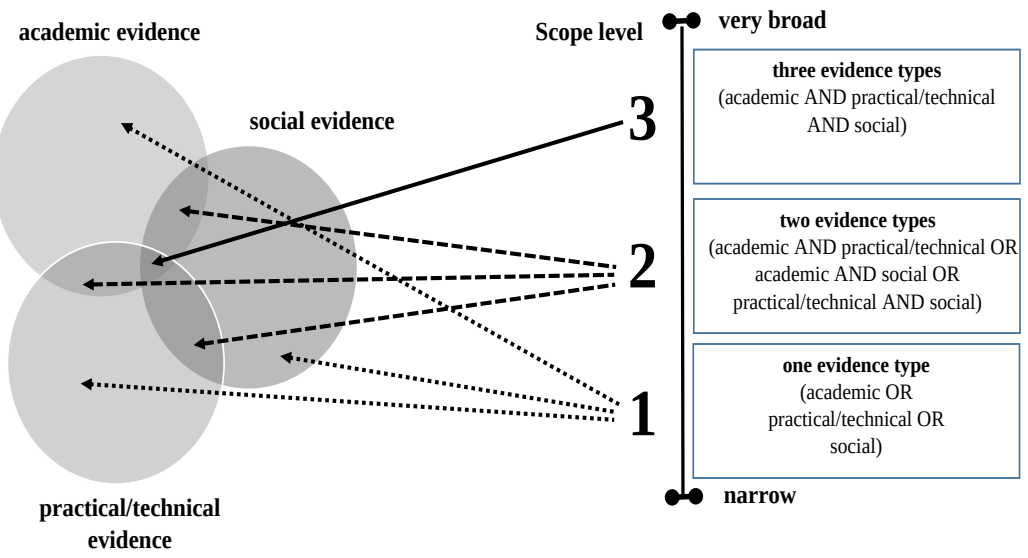
Dimension	Description	Levels
Scope	Number of sources used; with the main categories: • academic evidence • practical/technical evidence • social evidence	1. one main type (= narrow) 2. two main types 3. three main types (= broad)
Strength and quality	Quality and trustworthiness of the available sources	1. soft 2. medium 3. hard
Evidence-orientation	Extent of using evidence in policymaking	1. opinion-based 2. evidence-supported 3. evidence-based

Source: Author

The scope of evidence is defined by the endpoints "narrow" (= scope level 1) and "very broad" (= scope level 3) – see Figure 3. Areas that do not overlap refer to settings in which decision-makers are only interested in special types of evidence. The intersecting areas present the case that more than one type of evidence is considered. It is possible that decision-makers consider the evidence provided both by researchers and implementing

organisations. In addition, the evidence provided can be the consequence of a joint collaboration between academics and practitioners.

Figure 3: The scope dimension: types of evidence



Source: Author

The broad categories consist of sub-categories. The provided academic evidence can rest only on the results of one discipline, for example economics. It is also possible that – regarding very important questions, for example sustainability – decision-makers are interested in broad-based academic evidence from many disciplines. Depending on the question, the type of activity/intervention and the purpose, there is probably a need for different types and sources of evidence (Guijt & Roche, 2014). This is important for cases in which the assessment of challenges or relevant options differ across disciplines. However, because of the greater levels of complexity and the additional time needed for comparing different evidence, decisions are probably more difficult and therefore often delayed.[1]

The internal and external decision-making contexts affect what constitutes evidence and how that evidence is utilised (Dobrow, Goel, & Upshur, 2004).

(2) The strength and quality of evidence

The lateral axis of the evidence cube shows the strength and quality of evidence. At the core, it refers to the quality and trustworthiness of the available sources as used in ranking schemes (Table 1) originally developed in EBM, but in the meantime also used in other policy fields, for example environmental management (Dicks, Walsh, & Sutherland, 2014). Consultative techniques, expert opinions without explicit appraisal, etc., represent the

1 A fully rational decision-maker would compare the overall benefits and costs from including several pieces and types of evidence. According to the economics of evidence, they would extend the volume of evidence acting as input in decision-making up to that quantity in which the marginal costs (e.g. delay in decision-making, administrative costs, etc.) are equal to the marginal benefits (e.g. benefits also for marginalised people). The corresponding volume of evidence maximises public welfare.

lowest level ("soft"). The highest level ("hard") includes systematic reviews based on experimental or quasi-experimental evidence. Table 1 shows that systematic reviews are not always best. In the case of poor quality, for example if there are only a small number of observations, the level is downgraded. Much research is also flawed by unclear objectives, poor design, methodological weaknesses, inadequate statistical reporting and analysis, the selective use of data and conclusions that are not supported by the data provided. This is also covered by the categories of "strength" and "quality".

Table 1: Quality of evidence in the case of EBM

Level	Source type	Level
1	Expert opinion without explicit critical appraisal	soft ↕
2	Case series (a poor-quality cohort and case-control studies)	
3a	Individual case-control study	
3b	Systematic reviews (with homogeneity) of case-control studies	
4a	"Outcome" research, ecological studies	
4b	Individual cohort study (including low-quality randomised controlled trials – RCTs)	
4c	Systematic review (with homogeneity) of cohort studies	
5a	All-or-none studies	
5b	Individual RCT (with narrow confidence interval)	
5a	Systematic review (with homogeneity) of RCTs	hard

Sources: Based on Camfield, Duvendack and Palmer-Jones (2014) and OCEBM Levels of Evidence Working Group (2011)

Table 2 provides a metric based on Table 1, which enables the inclusion of sources beyond evidence.

Table 2: Quality of evidence in the extended framework in absolute terms

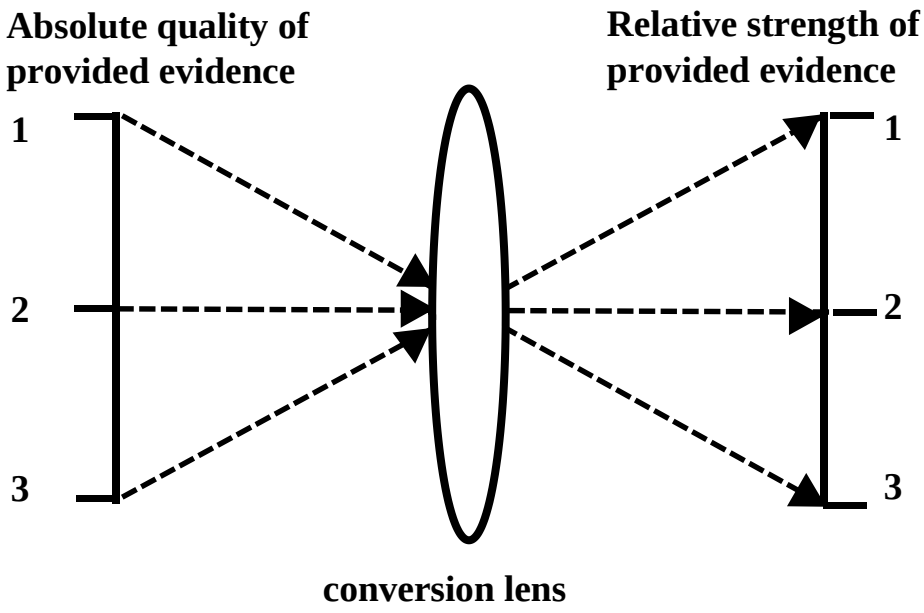
#	Source types (examples)	Level
1	Expert opinions with limited critical appraisal, results from case studies; data from results-based management and performance data with low quality; single and individual statements from representatives of interest groups	soft
2	Individual research outputs based on quality-based methods; results-based management and/or performance data with good quality; statements from citizens' movements	medium
3	Systematic reviews of research outputs based on rigorous methods; highest-quality results-based management and/or performance data; very broad-based public petitions and voices from civil society	hard

Source: Author

Although it is understood to be a general measurement tool for every kind of evidence, it is clear that the quality available in research science is not always achievable in other settings. For example, based on the insight that the quality of the scientific evidence used in legal processes is different from research science, Jasanoff (2006) argues that the former should never be subject to the kind of ongoing communal scrutiny. However, this is not the claim of Table 2.

The quality of evidence in absolute terms should not be confused with the *strength of evidence for policy-makers*. In absolute terms, all three types of evidence included in the scope dimension are inherently linked with different levels. Academic evidence provided will probably be of a higher absolute quality than social evidence (Figure 4 – left side). The relative strength of provided evidence (Figure 4, right side) is rather the consequence of how evidence is perceived by policy-makers. For example, soft evidence provided by qualified stakeholders who are understood as being very important will probably rise in the categories on their level. The corresponding process is illustrated in Figure 4 using the conversion lens.

Figure 4: Perception of evidence by policy-makers



Levels of absolute quality and relative strength

1 = soft
2 = medium
3 = hard

Source: Author

(3) The extent to which evidence is used in policy-making

The extent to which evidence is used in policy-making depends on many critical factors and is highly dependent on the setting.

In general, there are three different *process-related* steps for considering evidence:

1. Perception: Policy-makers become aware of available evidence.

2. Input: In the decision-making process, available evidence is theoretically taken into account, for example through commitments.

3. Implementation: Evidence is actually included in practical applications.

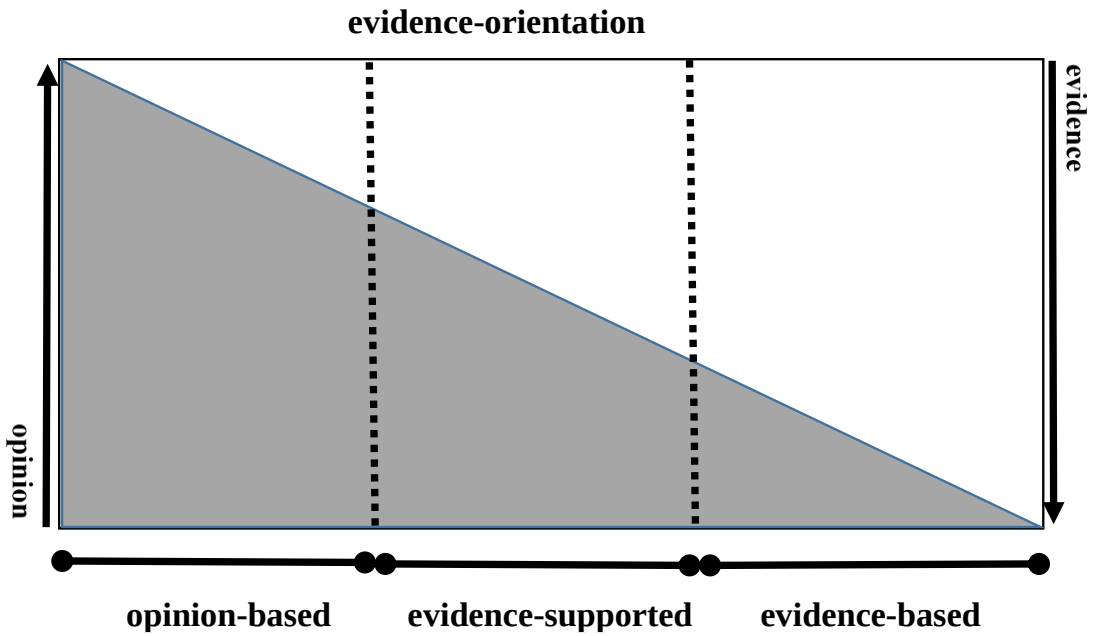
Figure 5 presents the level of evidence-orientation in policy-making. Using "opinion" and "evidence" as variables, three basic levels are distinguished:

- At a minimum level, policy-makers completely ignore available evidence. This is the case of *opinion-based* policy-making. Evidence is used, if at all, in a very selective way, mostly for justifying ex post their own views of the world. This level can also be characterised by the predominance of ideological standpoints and prejudices. Opinions also dominate in cases in which there is no evidence available because of the novelty of the problem and reasonable doubts about the provided evidence.

- At the other extreme, *evidence-based* policy-making presents a situation in which decisions are only based on evidence, for example in the case of a very pragmatic decision-maker for whom ideological aspects do not matter at all. This stage also corresponds to more benevolent decision-makers who are results-oriented. Information about what works can improve the outputs, outcomes and impacts of interventions and activities. EBPM also refers to more rational-oriented approaches in decision-making, in which decision-makers screen the available evidence in their attempt to compare action alternatives. The distance between the lower and upper ends of the spectrum corresponds to a situation in which a decider considers ex ante many different components, including some kind of available evidence. There might be many reasons why evidence is used only in a limited way. In a situation dominated by political features, available evidence might only be used for supporting existing political priorities.

- *Evidence-supported* policy-making can also be the consequence of a poor understanding of the value of evidence or limited trust in its reliability. At this stage, opinions can still be relevant.

Figure 5: Evidence-orientation: the extent to which evidence is used in policy-making



Source: Author

Theoretically, a situation in which all elements reach their highest value could be desirable. Realistically, it is frequently either not accessible at all or the costs involved for reaching that point would be too high. Collecting available evidence is time-consuming (search costs) and related to other (opportunity) costs. Therefore, the economics of evidence matter.

The corresponding evidence must be processed and prepared for becoming a menu for decision-making. For example, it is not enough if academic evidence is published in academic journals, because policy-makers probably do not read them. There is a need for think tanks and brokers to translate, aggregate and synthesise available evidence in order for it to be seen. In addition, not all sources of evidence are sufficiently sound to form a basis for policy-making. In the case of very poor data and information, a critical discussion is needed to include or exclude them as an option in the decision-making process.

The advocacy process is not linear. There is a need for feedback loops. Selecting the appropriate evidence can also be understood as a form of co-production between policy-makers, advisors and some other stakeholders: Knowledge-making and decision-making in such systems are continually reshaping one another (Clark, van Kerkhoff, Lebel, & Gallopin, 2016).

3 Providing evidence in development cooperation

3.1 Sources of evidence

Providing evidence in DC is challenging for many reasons. Often, the case of EBM is understood as a blueprint. Consequently, the EBM concept is immediately transferred to the DC sphere without considering that the latter context is much more complex. The field of human health is the focus of only one discipline (medicine). There are different concepts of health, including modern and traditional medicine. However, in selecting reasonable candidates for intervention, evidence based on different health concepts is not synthesised. In contrast, DC is the research area of many more (sub-)disciplines working on a different understanding and interpretations of reality. Therefore, evidence concepts in DC can also differ. In some settings, it is particularly important to combine the available evidence. DC provides a rich basis for academic, practical-technical and social evidence. Below is a presentation of the most important types of sources.

A broad spectrum of sources and instruments can be used for generating and disseminating available evidence. Unstructured data forms the foundation of evidence. However, the value of data is limited before being processed and analysed. Special formats such as academic journals, reports, papers and newsletters are used for providing and channelling academic, practical-technical and social evidence. Moreover, the corresponding stakeholders and agents often collaborate closely in generating evidence, for example in the case of projects. There are also platforms for knowledge-exchange. Sometimes staff members of practical-technical organisations publish results achieved in academic journals. Researchers also collect quantitative and qualitative data and information, which are important resources for providing evidence. The results are often disseminated as articles and books. Frequently, the creation of data and information is a by-product of DC, for example in technical project reports. Governmental bodies, international organisations – and increasingly, non-governmental providers of DC – collect monitoring data for mapping the performance of programmes and projects. At the partner-country level, statistical data (e.g. census data) are collected. Household surveys are an important source of evidence. Recently, the collection of management data – and, in particular, evaluation results – has become more important. Those data and information are often restricted in the organisational domain. However, summary reports are often available for reasons of accountability. Implementing agencies also provide technical reports that reflect the project results. In DC, civil society is of particular importance. A common understanding is that people's needs should be adequately mapped in policy-making. Civil society organisations collect social evidence and disseminate it through mass media, public events and/or political bodies. According to the concrete country-setting, informal socio-cultural evidence is used additionally as an input in decision-making.

Data for development is a topic that has gained a lot of momentum. On the one hand, there is the discussion about the existing data gap for covering the SDG targets, in particular in the case of partner countries and some settings. For example, there is a lack of evidence for implementing sector programmes. On the other hand, the volume of big data is expanding, and there is the question of how to utilise the corresponding information. The number of sources and corresponding materials for evidence-creation is still exploding. These include widely used technologies such as mobile phones; global positioning systems and geographical information systems; social media; web-based databases; audio and interactive

voice response; and digital videos and cameras. Meanwhile, much information is also accessible through the internet without existing technical and financial barriers.

The following presentation has a particular focus on instruments that have recently become the focus of attention.

3.2 Statistics and monitoring

For many years, different types of data have been collected at the international, national, regional and local levels by international organisations, national governments, line ministries and administrations; by DC providers and their implementing agencies; and also by researchers. Indicators, which are a special category of data, are highly relevant in DC. They provide crucial information for different stakeholder and are used for planning, managing, monitoring and evaluating (Holzapfel, 2014).

In operational terms, evidence corresponds with effectiveness, which is the relationship between planned and achieved results. Performance and results indicators are important points of reference for providing operational evidence about doing the right things. Therefore, monitoring is used. It is a continuous process of regular, systematic assessment based on participation, reflection, feedback, data collection, analysis of actual performance using indicators, and regular reporting in the life of a programme or project. The purpose is to track project and programme progress towards – and achievement of – results and quality. Results from monitoring can help shape eventual corrective actions for policies, programmes and practices that need improvement.

The MDGs heightened interest in the role of indicators for measuring global development progress and helping to underpin evidence-oriented decision-making. In this regard, household surveys are a rich source of information. However, since the introduction of the MDGs, severe shortcomings in the process of the collection and provision of data have become clear. There are still huge problems regarding the collecting of data in partner countries. In many cases, basic information is simply unavailable. For example, as many as 100 countries still do not have accurate records of births and deaths. The SDGs also include many areas in which governments have never previously attempted to keep accurate data. In addition, in terms of some SDG targets, there were problems in finding adequate indicators. Therefore, several initiatives for strengthening data for development were started (Keijzer & Klingebiel, 2017; OECD, 2017).

3.3 Results management

The desire to achieve results has always been on the DC agenda. Historically, the focus on results can be seen as part of the new public management paradigm, modelled on corporate-sector practices designed to maximise shareholder profits and eschew any explicit ideological commitments (Eyben, 2015). In the 1980s and 1990s, results management was already being applied in project aid management, supported through methods such as the Logical Framework Approach (Vähämäki, Schmidt, & Molander, 2011). DC is undergoing a radical change, and the reality of the aid landscape has shifted. There are problems related to fragmented aid (Klingebiel, Mahn, & Negre, 2016), and there is recognition that the

existing evidence base on the effects of development interventions has been either weak (Savedoff, Levine, & Birdsall, 2006) or unclear, as in the case of the "micro-macro paradox" of aid (McGillivray, Feeny, Hermes, & Lensink, 2005). Since the turn of the millennium, there is a growing interest in the question of how to arrange development aid so that it is more effective (Klingebiel, 2014). These issues have challenged donors and partners to improve the results of DC activities, despite the recognition that the aid dependency of partner countries has decreased over time. In addition, the ongoing call for justifying aid expenditures has created an increased demand for more measureable and credible results for presenting the most concrete evidence (Klingebiel, 2012). However, the relevance of the aid effectiveness agenda is also questionable because of the diminishing overall importance of aid for partner countries.

Therefore, "new" results-oriented models were adopted that reversed the intervention logics that were being used. These concepts built on previous work beyond DC, for example the 1993 US Government Performance and Results Act. Until that time, the guiding principle had focussed on the performance of project implementation, meaning tracking inputs, activities and processes. With results-oriented concepts, there was a shift towards monitoring results, in which outputs and outcomes are used as analytical starting points for studying the results chain. Defined appropriately, they are measured using a set of numerical indicators. The new guiding principle also aimed at defining the milestones being achieved, quantified by numerical targets. In this context, results management was transferred to the programme and organisational level. Significant differences to former management approaches also exist in terms of the used terminology.

Results-based management (RBM) was defined by Organisation for Economic Co-operation and Development/Development Assistance Committee (OECD/DAC, 2002) as "[a] management strategy focussing on performance and achievement of outputs, outcomes and impacts". It is a management concept. Three years later, the OECD/DAC adopted the term "management for development results" for incorporating new ideas about collaboration, partnership, country ownership, harmonisation and alignment, and by providing a higher management standard by asking stakeholders to continuously focus on country outcome performance rather than short-term results.

In the Paris Declaration on Aid Effectiveness (2005), the term "management for results" is used to denote the fourth principle of the Declaration, under which "results-based management" and "results-oriented reporting and assessment frameworks" are highlighted as means towards achievement.

The implementation of results management differs from country to country due to different priorities and strategies. Results or outcome frameworks are links between country-strategic goals' higher-level outcomes, country organisational structures, key stakeholder relationships and development partners. They show the hierarchy of key outcomes for sector or overarching programmes and may include multiple layers – at the sectoral, regional, agency or individual level. They are used to capture results-based indicators so as to provide operational evidence of the achievement of narrowly defined results (Roberts & Khattri, 2012). The results matrix is an important building bloc in results-oriented concepts. It provides a good basis to support meaningful analysis throughout the results chain: from activities to impact.

German DC is using a results matrix that provides the basis for GIZ and KfW programming and reporting. It draws on indicators and data from partner countries, but also own data is collected. According to the 2016 report of the Global Partnership for Effective Development Co-operation, a multi-stakeholder platform to advance the effectiveness of development efforts by all actors, only 54.7 per cent of results indicators were drawn from country-led results frameworks (Global Partnership for Effective Development Co-operation, 2016). Quantitative and qualitative indicators are used to examine how well a project, programme or policy is performing against expected targets. Evidence is created through a continual process of feedback, learning and adjustment, which is supported through internal audits and management reviews, performance monitoring, iterative programme implementation and evaluations (OECD, 2015).

In the case of the United Kingdom, DFID has developed a results system that includes all the essential elements needed. It includes a policy and management framework; targets and indicators; and the capacity and incentives to collect, process and analyse results information. The framework, which is similar to models adopted by major multilateral development banks, enables the monitoring and management of the progress of development results at the corporate level, as well as the ability to publicly report on delivery. The system rests on general indicators for corporate reporting; more specific indicators are used for project reporting. Mostly, the system refers to output indicators (OECD, 2014b) and consists of several levels: Level 1 measures progress on key development outcomes that DFID is seeking to contribute to in partner countries; Level 2 includes indicators that measure outputs and intermediate outcomes that can be directly linked to DFID interventions; Level 3 covers the operational effectiveness. Corresponding improvements can lead to better delivery of results and a better value for money. At Level 4, indicators focus on measuring organisational effectiveness. The corresponding indicators aim to monitor improvements in internal corporate processes (OECD, 2014b).

DANIDA introduced a framework for results management and reporting at the country level. It consists of a country policy paper, a country programme document and development engagement with each country. The framework is arranged around strategic and thematic objectives, engagement outputs and outcomes (OECD, 2014a).

Some authors argue that the claim behind RBM failed because of severe difficulties in handling the complexities of DC at the organisational level. Problems also exist because of the diversity of countries, sectors and contexts in which providers work. In addition, difficulties persist because a common understanding of what RBM is – and for what purpose it should be applied – is missing. As RBM is also based in management thinking with a focus on control and process, there is some clumsiness in adapting to changes. Another reproach is that RBM used by DC providers mostly focusses on donor priorities and not on the national priorities of partner countries. However, RBM is an important tool for providing operational evidence (Vähämäki et al., 2011).

3.4 Results-based approaches

Results-based approaches is an umbrella term for a variety of approaches that have recently been introduced. These approaches aim to shift the paradigm of aid from a focus on inputs and activities to a focus on results by reshaping the relationship between development

partners/donors and partner countries. Therefore, quantifiable and measurable results must be identified that are linked as closely as possible to the effects of DC (Klingebiel & Janus, 2014).

In general, these approaches can be defined as any programme where the principal sets financial or other incentives for an agent to deliver predefined outputs or outcomes and rewards for the achievement of these results upon verification (Musgrove, 2011). In the case of DC, a national or sub-national government body of a partner country normally acts as a principal. The contract partner or agent is usually an implementing agency or an individual (Grittner, 2013).

There are demand-side and supply-side approaches. In the case of results-based aid (RBA), funds are disbursed from one government to another. If individuals, households, public facilities, communities, non-profit entities or private businesses receive payments for some deliverables, the term results-based financing (RBF) is used. Beyond that general classification, plenty of approaches with certain features exist that were designed by aid agencies (Grittner, 2013). They vary widely and are related to the incentives and target groups as well as to the concrete level of the results chain they operate. According to Perakis and Savedoff (2015), some examples are:

- The immunisation services support (ISS) provided by GAVI helps countries to expand routine immunisation coverage – based on performance payments calculated from immunisation data provided by countries' administrative reporting systems.

- The Amazonas Fund aims to promote forest conservation; official development assistance (ODA) from Norway and Germany to Brazil is delivered on the basis of verified reductions in carbon emissions from slowing the rate of deforestation.

- Another example is the payment for secondary education in Ethiopia by the United Kingsom's DFID for students who take the exam.

The KfW has implemented several bilateral DC projects with an RBA mechanism on behalf of the German Federal Ministry for Economic Co-operation and Development (BMZ) within the framework of German financial cooperation, for example:

- a voucher programme in Kenya for medical treatment with subsidised prices or an output-based aid programme in Uganda's energy sector (Blume, 2012);

- the results-based financing for maternal and newborn health (RBF4MNH) in Malawi, led by the Reproductive Health Directorate of the Malawian Ministry of Health, which aims at improving the quality of selected Emergency Obstetric and Newborn Care facilities via supply-side incentives, demand-side payments and investment in equipment and infrastructure (White-Kaba, 2017).

For providing evidence, performance indicators are used. They must be well-chosen so that there are incentives to motivate partners to provide results of good quality. In practice, problems with data arise if the results are not assignable as a consequence of external effects. In this case, additional evidence could help (Holzapfel & Janus, 2015). In recent years, many DC providers have started experiments with specific RBA models as pilots, which are being widely applied in the health and education sectors. However, there is only a limited experience with these in the agricultural sector (Janus & Holzapfel, 2016). Because of that,

the evidence is mostly only applicable in concrete settings. Results-based approaches are not a magic bullet in DC and are prone to disincentives (Paul et al., 2018), but in the case of being well-designed, the approach is useful. Compared to RBM, results-based approaches provide stronger incentives to refer to available evidence because payments are linked to indicators. Empirical evidence supporting RBA is still scarce. A recently published study by Bernal, Celhay and Martinez (2018) demonstrates, in the case of the Saludad Mesoamerica Initiative in El Salvador, better results compared to conventional aid. By now, there is some practical experience from the pilots available that is being used for model improvement (Kemper & Sieler, 2016).

3.5 Evaluation

There is a strong link between the evolution of the evaluation function and the growing awareness about evidence. In the 1950s, evaluations were originally introduced in US-based organisations, including the World Bank and the United States Agency for International Development (USAID), with the aim of measuring and comparing results. At that time, there was a strong focus on appraisal issues. However, evaluation was only a snapshot; processes were not considered at all. It was not until the 1980s, when evaluation became an important instrument for transparency and accountability, that international organisations started to look closely at the long-term impacts of DC. Meanwhile, there are many different types of evaluation used for different purposes (Segone, 2007).

Even though both evaluation and results-oriented approaches rest on results, each is associated with distinct functions. Whereas RBM has broad coverage, evaluation focusses on in-depth analysis. The former is performed routinely to track progress as a starting point for learning, the latter only occasionally. RBM only tries to answer what results were achieved. Evaluation has two main applications: first, for internal instrumental purposes to improve delivery of the current intervention or policy – therefore, process evaluation or performance evaluation is used; second, for determining if a policy or programme has worked. In this function, the evidence generated by the evaluation contributes to learning in specific policy areas. Evaluations can focus on single projects, programmes or on the organisational level. However, evaluations are also integrated in RBM – as a special phase for providing operational evidence. The concrete understanding about the purpose of evaluations versus RBM differs from agency to agency (Binnendijk, 2000).

Additional demand for evaluations in the new millennium was driven by the MDGs, by donors looking for guidance and by the imperatives of NGOs (Bamberger, Vaessen, & Raimondo, 2015).

On the project or programme level, different evaluation and assessment methods are used for supporting decisions, in particular priority-setting (Barr, Rinnert, Lloyd, Dunne, & Henttinen, 2016). Economic evaluations assess the overall value of a project, programme or policy, taking into account its costs. A cost-effectiveness analysis (CEA) is used for comparing costs and outcomes of different options, such as interventions. For many years, cost-benefit analyses have been applied in the public sector. A cost-benefit analysis (CBA) attempts to measure the positive or negative consequences of a project; outcomes are expressed as far as possible in monetary terms. A CBA assists decision-makers in various ways: first, to decide whether a proposed project or programme should be undertaken;

second, to decide whether an existing project or programme should be continued; third, as an instrument for choosing between alternative projects or programmes. The cost-benefit ratio (or benefit-cost ratio) provides information on whether or not intervention A has a better value-for-money compared to intervention B.

Recently, impact evaluations have attracted much interest. Unlike project or programme evaluations, which mainly concentrate on whether an intervention was generally successful, impact evaluations focus on whether a development activity had an effect on certain target groups. Therefore, the counterfactual is considered – these are outputs and outcomes in the absence of the intervention. Evidence provided by impact evaluations can be used for deciding whether an existing project should be scaled-up or stopped. It can help to improve the design of development projects, programmes or policies in order to compare different projects and programmes with regard to their effectiveness. It can be used for improvement and learning. However, the use of RCTs and robust impact analyses also has its limitations. For example, they cannot be implemented in programme-based aid and in the context of institution-building.

There is a broad consensus among researchers and practitioners in DC that impact evaluations should be used for analysing the impact of interventions (Rudolph, von Schiller, & Strupat, 2017). Because of this, the number of evaluations carried out in the development sector has increased enormously. According to the International Initiative for Impact Evaluation (3ie), fewer than 20 studies per year were published before 2004. The number increased dramatically after 2008. In the content of a web-based repository, Cameron, Mishra and Brown (2016) identified 2,259 studies published between 1981 and 2012. Whereas prior to 2000, 81.1 per cent of the studies were published in health journals, impact evaluations in sectors outside of health became more popular thereafter (Cameron et al., 2016). Through the years, the variety of types and approaches also grew.

Impact evaluation approaches aim to demonstrate that development programmes lead to results. In analytical terms this means that the intervention as cause has any effect. There are intended and unintended causal changes, as well as positive and negative ones. Even if a project or programme is effective in terms of meeting its goals, it might not be judged as a success if it has also generated large negative side effects. The definition of impact determines the scope and content of the corresponding study. The main purposes for impact evaluations are: 1) advocacy (value of an intervention for informing future policy design); 2) allocation (helping to prioritise which projects, people and institutions are given funding); 3) analysis (learning about improvements and what is working) and 4) accountability (as required under legislation and better practice) (Rogers, Hawkins, McDonald, Macfarlan, & Milne, 2015). Different concepts of causality (what produces the impact) and different approaches to "causal inference" do this in different ways. There is a need for different evaluation designs to cover the particular fundamental logic (Stern et al., 2012). Each of the design approaches included in Table 3 has strengths and weaknesses that might have consequences on the quality – and therefore on the kind – of identified evidence. For example, statistical and econometric models can have difficulties with multiple causalities and struggle to capture the interactions among variables or represent irregular, complex paths. Overall, these approaches help to answer key questions for EBPM (Donaldson, Christie, & Mark, 2009).

Table 3: Main approaches of impact evaluation		
Design approach	**Specific variants**	**Basis for causal inference**
Experimental	RCTs, quasi-experiments, natural experiments	Counterfactuals, the co-presence of causes and effects
Statistical	Statistical modelling, longitudinal studies, econometrics	Correlation between cause and effect or between variables, influence of (usually) multiple isolatable causes on a single effect, control for confounders
Theory-based	*Causal process designs*: Theory of change, contribution analysis, impact pathways *Causal mechanism designs*: Realist evaluation, congruence analysis	Identification/confirmation of causal processes or chains Supporting factors and mechanisms at work in context
Case-based	*Interpretative*: Naturalistic, grounded theory, ethnography *Structured*: Configuration, process tracing, congruence analysis, qualitative comparative analysis, within-case analysis, simulations and network analysis	Comparison across and within cases of combinations of causal factors Analytic generalisation based on theory
Participatory	*Normative designs*: Participatory or democratic evaluation, empowerment evaluation *Agency design*: Learning by doing, policy dialogue, collaborative action research	Validation by participants that their actions and experienced effects are caused by the programme Adoption, customisation and commitment to a goal
Review and synthesis	Meta-analysis, narrative synthesis, realist synthesis	Accumulation and aggregation within a number of perspectives (statistical, theory-based, ethnographic, etc.)
Source: Stern et al. (2012)		

Even though RCTs are often understood to be the "gold standard" of evaluation (Bédécarrats, Guérin, & Rouband, 2017), their results and relevance must be interpreted with care (Vaessen, Raimondo, & Bamberger, 2015). For example, the counterfactual answers only setting-specific questions (e.g. did it work here, for this particular group?) and cannot be generalised to fit other settings (low external validity). The design analyses only linkages between intervention outputs (causes) and outcomes (effects) and does not examine processes (what happens between intervention outputs and outcomes): It does not explain how the outcomes are achieved, or how and why the assumed causes contributed to the outcomes (Bamberger, 2015).

In recent years, the quantity and quality of available evaluations have improved a lot due to the increasing efforts being undertaken by many countries to reform evaluation systems. Some new public and private institutes have been established. For example, in the United Kingdom, the Independent Commission for Aid Impact (ICAI) was created in 2011. In the same year, BMZ established the German Institute for Development Evaluation (DEval). There are now many evaluation associations, societies and networks worldwide that provide capacity-building (OECD, 2016a). Also NGOs have developed strong evaluation capacities

and established their own networks that share among themselves and link local, national, regional and international entities (Segone & Ocampo, 2006). In 2008, 3ie was created, which provides a very useful website[2] with a comprehensive package of resources.

In Germany, the main providers of DC evaluation are the BMZ Evaluation and Research Division; DEval; GIZ Corporate Evaluation Unit; and the Evaluation Department of KfW Development Bank. Already in the early 1970s, BMZ had started to develop a sophisticated system to evaluate development policy. Therefore, in 1971 an inspection unit – and later on an evaluation unit (*Zentrale Erfolgskontrolle*) – were established (Wollmann, 2017). In the meantime, the BMZ's evaluation unit has acted mostly as a guide for the overall system and delegates the implementation of evaluations. DEval provides independent and strategically relevant evaluations for all German development activities and GIZ performs strategic and corporate evaluations. GIZ conducts about 100 project evaluations per year, but also produces studies on specific themes to learn from experiences across a range of programmes, drawing on findings from various evaluations, reviews and progress reports. GIZ also introduced a management response for strengthening follow-up on recommendations from evaluations. KfW focusses mostly on project-specific standard ex post and impact evaluations. KfW also carries out the analysis of specific themes using the data available from its ex post evaluation reports. Both GIZ and KfW have introduced several mechanisms for strengthening the use of evaluation-based knowledge related to future planning and decision-making. Both implementing agencies also created incentives for reinforcing the evaluation culture within their respective organisations (OECD, 2015). In contrast to the United States, in which Michael Kremer and Esther Duflo implemented many RCTs in cooperation with NGOs (see e.g. Banerjee & Duflo, 2011), the collaboration with NGOs in Germany is rather limited.

Briefly: Evaluations have become one of the main accepted instruments to tell success apart from failure and to foster learning from experiences.

3.6 Review and synthesis

Single studies are frequently only of limited interest for policy-makers – especially if the sample size is low. Systematic reviews pool together the findings from many studies and can be based on different study types, including evaluations (Waddington et al., 2012). Because they address questions of whether and how interventions work across settings, they provide the strongest basis for policy-making and include such methods as narrative review, meta-ethnography, realist synthesis and systematic review. The growing number of review and synthesis studies, in particular systematic review studies, has provided a new impetus for so-called repositories of knowledge or sending-order, evidence-producing organisations. Examples of such bodies are the Cochrane Collaboration, the Campbell Collaboration, the Evidence for Policy and Practice Information and Coordinating Centre, and, particularly relevant for development, 3ie (White, 2014).

EBM is frequently used as a reference case for the usefulness of evidence in decision-making. Methods used in medicine and natural science, such as RCTs, are regarded as blueprints for providing evidence in other disciplines. In the hierarchical system of classifying evidence used by EBM, the creation of high-quality systematic reviews is the

2 See http://www.3ieimpact.org

leading method. However, in the area of DC, there are many problems related to synthesising evidence, mainly because existing studies in DC differ much more in their focus and approach, their methods, the included variables, and the quality of data and results. Because systematic reviews in DC are normally based on only a small number of included studies, they have much less power (Mallett, Hagen-Zanker, Slater, & Duvendack, 2012).

3.7 Interim conclusions

Data is *the* essential input for generating evidence. In this regard, there are still enormous gaps in covering the SDGs, in particular in low-income countries (LICs). Data availability differs also in terms of policy areas.

Until recently, convincing evidence about DC activities was only available to a certain extent for particular settings, mostly at the project level in some sectors, such as health and education. Monitoring data and information were collected for tracking and ex post assessment. More comprehensive results-management systems often performed poorly. Since the turn of the millennium, data and information availability has increased a lot. Monitoring systems provide a comprehensive set of performance data. However, in particular in partner countries, monitoring systems are far from perfect. RBM and results-oriented approaches now provide a sound base for operational evidence. Different types of evaluations are helping to make different categories of analytical and operational evidence available.

4 Use of evidence in development cooperation

4.1 Perception

After the turn of the millennium, there was still a strong disapproval of – and scepticism about – using impact evaluations in DC (Faust, 2017). At that time, there was also no actual demand from policy-makers for more and better evidence for supporting decisions. Thus, available evidence was only of limited relevance for decision-makers. Since then, the level of awareness about the value of evidence has increased among providers of DC.

Several factors were responsible for pushing evidence onto the agendas of OECD/DAC member countries. First, in some member countries, there was a stronger awareness of the value of the evidence already in place. In the United Kingdom, the political opinion towards the relevance of results measurement and management had changed before in a striking way. At the end of the 1990s, the Tony Blair government popularised the term "evidence-based policy-making" (Banks, 2009). The government clearly defined the results it aimed to achieve with its DC (Department for International Development, 2011). From the demand side (policy-makers), there was also a push to use more evidence, to demonstrate results and to put more importance on evaluations. Even if it was motivated above all by the strategic thinking of policy-makers to use evaluations as instruments for justifying decisions already taken (Stewart & Smith, 2015), the revised approach supported a cultural change: The main focal point in meetings with partners switched from talking about money spent to results and impacts. The United Kingdom also established the ICAI for reporting to parliament and

to review its ODA (OECD, 2014b). For several years, results and accountability have also been a major focus of US development cooperation. President Barack Obama's Government Performance and Results Modernization Act of 2010 was the foundation for a stronger focus on results (OECD, 2016b).

The delegates and representatives of the pioneering countries were agenda-setters on international platforms, supported by the activities of their implementing bodies and institutes. Recognition of the serious loopholes in DC became the push for establishing 3ie. Also, bilateral donors such as DFID and USAID followed by implementing policies and directing attention to impact evaluations (Levine & Savedoff, 2016). In particular, the DFID funded many RCTs and also influenced international debates enormously. However, it is not clear whether the corresponding results remain in the research domain or are also being used to improve their own programmes. In contrast, the Netherlands arrange a call for tenders before their own projects are implemented, and so the evaluation results are of benefit within the implementation. The Development Research Group evaluates the impact of selected World Bank projects and those of other multilateral development banks based on their own Strategic Impact Evaluation Fund.

International organisations and platforms, including the OECD, serve as important platforms for discussions about the "results agenda" and its elements. OECD products such as DAC peer review reports reflect the need for strengthening donor countries' systems, processes and activities in this regard.

Strengthening evidence perception is also a very important issue in partner countries. Therefore, some DC providers, organisations and networks have started programmes related to awareness-raising and capacity-building among decision-makers (Stewart, 2015).

3ie is working closely together with policy-makers in developing countries. To strengthen a "culture of evidence-based policy-making", they have organised "Demand Generation Workshops". In these capacity-building training programmes, presenters explain what sort of evidence is available through impact evaluations, how it can inform policy and programme design, and why this is important. These events help to increase awareness and skills with regards to the value of evidence in policy-making. In addition, "Impact Evaluation Clinics" are organised around "policy windows" to support developing-country agencies to implement evaluations in terms of concrete programmes (Monaghan, 2013).

In a similar fashion, DFID designed the Building Capacity to Use Research Evidence (BCURE) programme[3] in 2013 to address this issue across 12 countries. Over the past four years, BCURE has promoted the use of evidence by decision-makers, which has been an important contribution for improving development outcomes. Table 4 presents a framework that includes three categories showing the types of evidence used by corresponding policy-makers at three study sites (Rinnert & Brower, 2017).

The Africa Evidence Network[4] consists of researchers, practitioners and policy-makers from universities, governmental bodies and NGOs who work in Africa and are interested in the provision and use of evidence in decision-making.

3 See https://bcureglobal.wordpress.com/

4 See http://www.africaevidencenetwork.org

The importance of considering evidence on health and social issues is also recognised by BMZ. Since 2004 the Health Practice Collection, a joint initiative between GIZ and KfW, has been publishing case studies and evidence briefs, written by experts, including results that German-supported programmes generate during their implementation.[5]

Table 4: BCURE value of evidence use framework

	Transparent use	**Embedded use**	**Instrumental use**
Description	Increased **understanding and transparent use of (bodies of) evidence** by policy-makers	No direct action is taken as a result of the evidence, but **use of evidence becomes embedded** in processes, systems and working culture	Knowledge from robust **evidence is used directly to inform policy** or programme
Examples	***BCURE VakaYiko:*** Several roundtables were held to help bridge the gap between research and policy-making on climate change in Kenya and to help decision-makers acknowledge the full body of evidence on climate change in the country	***BCURE Harvard:*** The researchers worked directly with government technicians to create a Report Dashboard, designed to serve as a one-stop shop for over 50 indicators deemed crucial for evaluating the Mahatma Gandhi National Rural Employment Act	***BCURE University of Johannesburg:*** In South Africa the evidence map, published by the Department of Planning, Monitoring and Evaluation, fed directly into the decision-making of the White Paper on Human Settlements
Scope: The array of policy-makers impacted by the reform – is its impact far-reaching across actors?	+++ Intergovernment, policy teams and country offices	+ One local government ministry	+++ National-level policy
Depth: Impact of change, how large is the size of the reform? Is there a substantial change from previous practice?	+ No in-depth change in practice that would be directly attributable to BCURE, but a contribution to a set of follow-up actions	++ Evidence tool created and saw immediate use, 150,000 hits in the first year	++ The Human Settlements Policy is potentially reaching a large proportion of the population, however, overall effect has yet to be determined based on monitoring and evaluation results
Sustainability: How sustainable is the change in the use of evidence?	+ One-off meetings but with potential to influence further changes in the use of evidence	++ Evidence suggests this will be a prolonged change	++ Evidence used for several policy decisions with potential to influence further policy choices

Source: Adapted from Rinnert and Brower (2017)

5 See http://health.bmz.de/ghpc/index.html

4.2 Input into decision-making

Evidence uptake in decision-making can have an added value compared to opinion-based decisions if two critical assumptions are met. First, the quality and strength of the available evidence must be sound. Second, the volume of evidence related to a problem or setting needs to be comprehensive. If evidence is only available for some particular settings, evidence will probably be only of minor relevance. Until recently, this particular aspect was predominant. For example, in 2005, the OECD/DAC peer review about the individual development cooperation efforts of Germany noted that Germany's ability to track and report meaningful information was weak (OECD, 2006). In 2006 the authors of the report "When Will We Ever Learn?", published by the Evaluation Gap Working Group and which was initiated by the Centre for Global Development, draws the conclusion that there was little or no sound evidence available for assessing whether development programmes actually work. Published evaluation reports showed well what money was being spent and what direct activities or services were being delivered, but not whether the provided services gave rise to real benefits for the target groups (Gaarder, 2014). There were also too few quality studies (Savedoff et al., 2006). The main results of a workshop organised by the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE) in 2008 focussing on rigorous impact analyses in DC were that the available methodological knowledge was limited, and the corresponding insights were rather scarce. However, things have changed and moved on.

Since that 2008 workshop, things have improved greatly. The number of performed impact analyses by German researchers and operative departments of implementing organisations, including international collaboration, has increased significantly. The overall evaluation gap is closing, even if it has not yet been closed. There now exists a huge body of high-quality, policy-relevant research and evidence on issues of DC because of the growing demand for – and funding of – such products.

Evidence matters! There are now numerous examples of the use of evidence in DC (Carden, 2009; Newman et al., 2014; Young, 2006). However, the concrete answer to the question about the relevance of evidence depends on the used definition of evidence. It is also often difficult to quantify directly in what way – and to which extent – that provided evidence has contributed to changes in policy or practice. Evidence is used in different ways by decision-makers, for example for analytical and/or operational purposes.

Results information is used on the intervention, country and corporate levels for accountability or as a management tool. Currently, many DAC members are putting a stronger focus on the former for presenting the idea that aid is actually working on the project/activity and programme levels. The UN system and most international development agencies now use some variants of RBM systems and frameworks, mostly on a more organisational and programme-wide level. Results information is now being widely used for accountability and communication.

In a recent OECD survey about results measuring and management that was directed to ministries and development agencies, "tracking progress" was cited most often as the first priority (Table 5). However, results information is also used for decision-making, mostly at the programme/organisational level. Information about outputs, outcomes or impacts of DC

interventions are applied to a lesser extent for providing analysis or explanations of performance (OECD, 2014a).

Table 5: How results information is used		
Level	**Most-cited first priority**	**Most-cited second priority**
Agency/corporate	For accountability to parliament/ legislature (11)	In external communication (11)
Country	To track progress (9)	For decision-making (7)
Sector/programme	To track progress (13)	For decision-making (8)
Project	To track progress (15)	For decision-making (11)
Source: OECD (2014a)		

According to a recently conducted DFID survey looking at the use of evidence across the department, 60 per cent of the staff agreed that the use of evidence in the department had increased over the past three years (United Kingdom, 2014).

In the United States, all country-development strategies have results frameworks that set out development objectives, intermediate and sub-intermediate results, and performance indicators. These results serve as the basis for project designs and evaluations.

In Germany, GIZ and KfW use results information routinely to improve programme management (Deutsche Gesellschaft für internationale Zusammenarbeit, 2013; KfW Development Bank, 2017).

The current use of available data and information generated by results management is dependent on their intended use and the performance indicators. In theory, RBM approaches are a powerful instrument. However, in practice, there are many problems that limit the information value of such systems. For example, in the 2013 OECD "Managing and Results Survey", many ministries and development agencies encountered challenges when using their results systems. Among other findings, 5 (13) out of 28 respondents stated 100 per cent (often) "incomplete frameworks (lack of baselines or targets, unclear results)"; 3 (15) out of 28 respondents stated 100 per cent (often) "difficulties in linking budgets to results information"; and 3 (15) out of 28 respondents stated 100 per cent (often) "difficulties in selecting appropriate indicators that measure results at the correct level". Those difficulties often arise because of difficulties in handling complexities.

Existing guidelines, institutional documents and manuals provide rich information about the use of evidence in policy-making. The existing obligatory German RBM has some shortcomings. Whereas country strategies set the objectives for overall German DC and specific programmes for each partner country, BMZ is currently working on measuring and reporting results more systematically.

According to the most recent OECD/DAC peer review report, BMZ was not able to make full use of results information and evidence from other sources for strategic planning and communication. The absence of a full-functioning RBM system in BMZ limits the extent to which Germany can define success and measure its overall performance in supporting partner countries' priorities.

In the absence of overall indicators and targets at the organisational level, it is unclear how BMZ is using results information from programme implementation and other sources of evidence to feed into strategic planning and public communication. In contrast, GIZ and KfW routinely use results information at the project and programme levels to improve the overall corresponding management. BMZ obliges GIZ and KfW to adopt the results matrix and implement monitoring. Both organisations have good monitoring systems in place, drawing primarily on indicators and data from partner countries' own systems, which limits the need for additional monitoring and reporting requirements. In addition, most technical cooperation programmes include a specific component for strengthening partner countries' monitoring and evaluation systems (OECD, 2015).

In the context of the project cycles, there is some space for considering evidence more explicitly. For example, USAID included an explicit statement in their programme cycle operational policy that decisions about where and how to invest foreign assistance resources should be based on analysis and conclusions supported by evidence (United States Agency for International Development, 2017). The BMZ's "Guidelines for Bilateral Financial and Technical Cooperation with Developing Countries" (German Federal Ministry for Economic Co-operation and Development, 2008) stipulate that the economic, socio-economic, gender-specific and ecological impacts should be considered in the selection of projects. Recently, a procedural reform was implemented at the level of BMZ. There is now a stronger focus both on providing more existing evidence in project proposals and presenting project alternatives. However, there is still some space for improvement.

Evidence provided through evaluations now has great value if it proves that an ongoing intervention is successful. For example, in the case of a DFID social cash transfer programme in Zambia, evidence was provided that this policy has a beneficial effect on poor people's lives. As a consequence, it was scaled-up (Barr et al., 2016). The joint results-based financing programme RBF4MNH for supporting maternal and newborn health by the Malawian Ministry of Health and KfW was rated a success (White-Kaba, 2017). The government of Malawi declared that it is willing to continue that programme after the German funding is ended. Therefore, the programme was included in the national health-sector strategy. However, there is insufficient information in the literature about the consequences of missing positive evidence or negative implications in terms of projects or programmes. There is probably a bias in the literature.

Partner countries also use evidence in priority-setting for health programmes relevant to LICs. For guiding the allocation of funds, evidence on the burden of disease and the cost-effectiveness of selected interventions, such as the global burden of disease and CEA, are used. In talking about the use of evidence, policy-makers rarely mention the use of particular literature reviews. (Ideas about) evidence are used more in a rhetorical way as a part of arguments for what should be prioritised. For many UN and bilateral agencies, quantitative evidence is also helpful for making (instrumental) decisions about priorities. In contrast, NGOs also emphasise the value of qualitative information. Mixed-method approaches are nowadays often understood as providing reasonable background. However, as bilateral agencies and national organisations are obliged to implement government priorities, priorities are assigned politically and evidence is used symbolically for supporting decisions already made by politicians (Kapiriri, Sinding, & Arnold, 2017). Sometimes, partner countries also recognise the importance and added value of evidence. For example, in the

case of Zimbabwe, the government asked for the ex post evaluating impact of a programme supporting AIDS orphans (KfW Development Bank, 2016).

4.3 Barriers and facilitators

On a conceptual and theoretical level, many stakeholders agree that there is a need to use available evidence in policy-making, including information, data, statistics, metrics and measurements. However, some essential barriers limit the use of evidence in practice. These topics are not specific issues in the field of DC (see e.g. Oliver, Innvær, Lorenc, Woodman, & Thomas, 2014).

1. Political economy

Traditionally, the link between research and policy has been viewed as a linear or rational process, whereby a set of research findings shift from the "research sphere" over to the "policy sphere" and then has some impact on policy-makers' decisions (Young, 2006). However, new evidence will probably not have a direct impact on many decisions. Arguments for rationality ignore the differences between the rhetoric and reality of practice, and the vested interests versus knowledge-creation in the aid sector (McNulty, 2012). Policy-makers are often political actors. Thus, values, political beliefs and interactions with the political system are relevant. In practice, evidence is only one among a number of factors at play in policy processes (Parkhurst, 2016).

2. Experiences versus evidence

There might also be objections by decision-makers or administrative staff members because of their general aversion to innovations or the assessment that the provided evidence does not resonate with their experiences.

3. Timeliness and window of opportunity

Sometimes, available evidence is not considered in decision-making because evidence is not available when it is needed. For example, Bamberger (2015) points out that the results are sometimes presented after decisions have been made or the report is published too early – even before decision-makers or the public have begun to focus on that issue. Similarly, in a survey conducted by the US Government Accountability Office about organisational performance and management issues, only around 15 per cent of USAID managers said that they agreed to a great extent that they had access to the performance information they needed to manage their operations or technical work (United States Government Accountability Office, 2017).

4. Openness towards evidence

Policy-makers signalling some kind of openness towards using evidence limits their scope for decision-making. In this case, there are special justification needs if evidence is not accepted in making decisions, for example in cases in which available evidence clashes with their own ideological beliefs.

5. Contradictory and inconsistent evidence

Policy-making is challenging because of uncertainty, complex settings and different views of the world. Those components and characteristics of reality define the political universe, and therefore the demand-side of evidence. Often, there are no simple solutions because academic results are different. This is due to researchers belonging to different "schools" and applying various research methods as well as researchers having different disciplinary backgrounds and shifting perspectives. There are various kinds of evidence that often clash. Contradictory and sometimes inconsistent evidence acts as a barrier for accepting it as input in decision-making. Beyond that, the provided practical-technical evidence can also differ according to the capacity of the administration. In addition, social evidence can also take various forms due to particular social movements and interests.

6. Right and wrong questions

There is evidence that a key issue affecting uptake is whether the research provided is perceived as a solution to an existing problem or not (Young, 2006). Some policy-makers do not believe that research questions are relevant for them (Fourie, 2017). In the case of complex evaluations, it is also conceivable that the wrong questions and irrelevant findings are presented that prove to be of no interest to policy-makers. Sometimes evaluations also have a narrow focus on the impact of programmes, or there are difficulties in attributing causality for complex programmes.

7. Presentation of results

Even in the case of impact evaluations, which were demanded for accountability and as an instrument for more effectively achieving development goals, the type and form of results presentations matter. They must be presented in a way that is more meaningful to policy-makers. As White (2014) points out, influencing policy is about both the product and the process. The product, that is, the study, should address the full range of evaluation questions of interest to the policy-maker. According to Stern et al. (2012), these questions are:

- To what extent can a specific (net) impact be attributed to the intervention?
- Did the intervention make a difference?
- How has the intervention made a difference?
- Will the intervention work elsewhere?

Policy-makers do not read academic journals. They are not interested in t-statistics, probability values or other ways of presenting statistical significance. They want to hear about the importance of the results and the corresponding costs. In addition, researchers want to play some part in ensuring that policy-makers are aware of these findings if they are to influence policy. The policy impact will be greater if there has been engagement with policy-makers and programme managers from the outset, starting with establishing the evaluation questions. However, policy-makers' perceptions of why they do or do not use evidence is not necessarily the same as the objective reality.

8. Value of evidence related to exiting policy frameworks

If the given recommendations – based on the results of a study, for example an evaluation – do not fit the policy framework of the contracting authority, evidence will probably not be directly considered. That was the case with a DEval study about general budget support in sub-Saharan Africa (Krisch, Schmitt, & Dörr, 2015). BMZ concluded that the existing policy framework of the government limits the use of some results.

9. Existing evidence culture

At the national level, the "evidence culture" – and the interests of policy-makers and the main implementing bodies for using evidence – is mixed. There is evidence that in countries such as the United Kingdom, it was possible to bring up an "evidence culture" – one that is more open to perceiving and using evidence. This means embedding the use of evidence into existing organisational structures and processes (Baker & Salib, 2017). Therefore, a political push or facilitators are needed. For example, there seems to be a stronger culture for using evidence derived from impact evaluations in Latin America as a consequence of the programme PROGRESA in Mexico. But there is also a growing interest in evidence-based policy in Africa, for example in Nigeria (Uneke et al., 2011).

With the aim of improving the adaptation of evidence into decision-making, the topic "knowledge translation from research to practice" has attracted much attention. Plenty of models and approaches have been developed about how policy-makers can be motivated to implement more available evidence in decision-making (Howlett & Craft, 2013; Young, Ashby, Boaz, & Grayson, 2002). In fact, the usefulness of the suggested approaches critically depends on the setting. The underlying concept of the following case study (Section 5) is just one (promising?) approach.

4.4 Interim conclusions

Recently, changes in attitudes of policy-makers have been observable. Some governments and international bodies have started to push EBPM on the agenda. This, in turn, has had an impact on the generation and supply of evidence.

There is evidence that the importance of evidence in DC has increased. The perceptions at the level of policy-makers in donor countries and – to a lesser extent – in partner countries have improved a lot. Practical-technical evidence is being used for operational purposes in projects and academic evidence for analytical purposes in guiding policy-making and the selection of intervention strategies. In contrast, social evidence is often not directly considered in practical decisions at the level of administrative bodies and governmental organisations. However, social evidence sometimes has a direct impact in the policy-making process, and its input can influence decisions.

Policy-makers are often political actors. Therefore, available evidence is frequently not used for implementing welfare-maximising policies, but rather as an instrument for justifying political statements, intentions and implementing policies that have already been decided on.

Whether or not evidence is considered is dependent on the setting of the decision problem and its environment. To improve the chances of evidence being considered in policy-making, barriers must be lowered. The introduction of an evidence-based culture and reasonable evidence-transmission concepts seem to be very important.

5 Case study: the Copenhagen Consensus

5.1 Background and approach

The Copenhagen Consensus Centre is an international Copenhagen-based think tank established in 2002.[6] Since 2004, several "projects" have been conducted that focus mostly on development problems at the regional, national and global levels, such as Andhra Pradesh, Bangladesh, Haiti or overseas development spending. Some projects focussed directly on the world's biggest problems (Lomborg, 2007b) or the post-2015 Development Agenda (Lomborg, 2015); others paid particular attention to thematic issues, for example HIV (Lomborg, 2012) or climate change (Lomborg, 2007a). All of them use academic evidence for priority-setting at the very least. Because of this and some particular characteristics (see below) of how the people involved discuss and communicate, the Copenhagen Consensus (CC) was chosen as a case study.

Based on the funding provided by governments and foundations, optimal solutions for the problems under review are identified by using the routine economic concept of CBA, a valuation technique that is also used for ex post evaluations. In a CBA, the net benefit of investments is calculated by comparing the expected returns from alternative policies (= value for money). The ultimate logic behind this is to choose the more effective ones. Thus, from a rational perspective, the CC approach can be regarded as an optimal approach for providing evidence for policy-making. Most of the projects were processed in a similar way based on the CC methodology.

5.2 Bangladesh priorities

For example, it was the aim of the "Bangladesh priorities" to identify smarter solutions for the existing challenges in that country against the background of limited resources and time. The whole project was a map exercise. The assumed annual budget line for the government was $30 billion, with $3 billion being given in development aid by outside organisations. Policy-makers, international donors, NGOs and businesses were addressed as the main stakeholders. The project actually lasted for more than a decade (Lomborg, 2017).

The Bangladeshi experiment started with an analysis of the country's seventh Five Year Plan as a focal point for discussing national development. All of the included 20 topic areas were adopted in the analysis. In collaboration with the Bangladesh Rural Advancement Committee – one of the largest development-oriented NGOs in the world – the Swedish International Development Cooperation Agency and the Danish embassy in Dhaka, more than 800 people from the Bangladesh government, academia, think tanks, NGOs, the private

6 See http://www.copenhagenconsensus.com/our-story

sector and development organisations were invited to articulate their recommendations. Finally, there were more than 1,000 ideas offered, including many topics that were also mentioned in the official plan, such as infrastructure, education and public health.

The ongoing research was presented in the largest Bangladeshi newspapers, and the proposals were ranked in Rural Village Forums and Youth Forums. Based on background studies, 20 roundtables with Bangladeshi experts were held to comment on the proposals and to assess them based on features such as potential, political support and available data. Based on that procedure, 76 promising proposals were identified. CBAs were used for assessing all of them. In the case of "wetland conservation in the Sundarbans", the expected benefits were calculated to be almost $4 billion for a cost of $1.4 billion, implying nearly $3 in benefits for every $1 spent. Most of the interventions scored between 1 and 20. For example, the score of an early childhood education programme was 18, meaning that every $1 invested would bring an expected $18 in return. However, there were certain interventions – including unconditional handouts of cash to poor Bangladeshis and immunizing against cervical cancer – given lower scores, down to 1. The best solution, digital procurement, was identified as having $663 in benefits for every $1 spent (The Economist, 2016). The point of reference for this recommended intervention were the enormous inefficiencies in government procurement, including corruption.

5.3 Appraisal

Past and ongoing CC projects have attracted much attention. The concept has strong academic appeal. Because of the product design and the corresponding public relation campaigns – including books in accessible languages, academic papers in high-quality journals (*Nature*, *The Lancet* etc.), magazine articles (*The Economist* etc.), and contributions in important newspapers – many academics and political stakeholders were able to see CC results. For many average people, the method is also easy to understand because the available options are ranked. CC projects also work with a high degree of transparency. Much background information is accessible on the Copenhagen Consensus Centre website. Experts in that field, sometimes Nobel laureates, present the results. For many people, the evidence is plausible. Overall, the CC approach seems to be a successful model for translating existing evidence into policy advice as an input for decision-making.

Even in the case of the CC, it is not possible to make a full appraisal of the success of that approach. Nevertheless, the CC model has some key features that are interesting in terms of knowledge transmission:

a) the involvement of some high-ranked experts in thematic issues. Some of them are Nobel laureates and so are winners of a prize with an excellent reputation. Even non-academics make associations between the Nobel Prize and academic expertise;

b) the use of a methodological approach that is well accepted in economics, routinely used in the public sector and that is easy for non-economists to understand;

c) the reappraisal of the relevant evidence and the corresponding transparent documentation of the assessment;

d) the inclusion of alternative views;

e) the hierarchical ranking of options according to their value for money;

f) an excellent marketing campaign, in which the recommendations are proofed by citing experts.

However, the CC concept is highly controversial.

First, it is a project dominated by economists. The invited scientists are mostly recruited from researchers working in that field. According to Bjorn Lomborg, head of the Copenhagen Consensus Centre, there is a "need for economists to set global priorities" (Lomborg, 2004a). Such a statement would not be acceptable for many scientists from other academic disciplines. In a world of pluralism, the world's priorities should not be defined by economic imperialism. In the words of Jeffrey Sachs: "Economists alone are not enough" (Sachs, 2004).

Second, many scientists acknowledge the academic value of the background reports, which are used as an important input for a CBA, but they disagree with some of the models and assumptions used in the CBA. For example, many academics working in the field of environmental sciences opposed the CC results in assessing climate change. The project made headlines for rejecting a policy framework for the control of climate in the context of global priorities (Lomborg, 2004b). Based on a more recently published CC analysis about smart solutions to climate change, Lomborg concluded that there was no need to reduce CO_2 emissions to any significant extent in the near future. More spending on green research and development would be more reasonable (Lomborg, 2010). One major point of the corresponding discussion with climate scientists was about the level of discounting future costs and benefits (Hamaide & Boland, 2006). This is a tool used by scientists to transfer future values into the present time. The concept rests on the assumption that people are inpatient and prefer present time. Because of the high discounting of future benefits of climate control measures, their present value is very small, so it does not seem to be a good investment because of a systematic undervaluation of the future (Guo, Hepburn, Tol, & Anthoff, 2006). Thus, climate scientists argue that delaying a response would worsen the impact of these challenges. It was also stated that the use of a marginal CBA is not adequate for addressing the problem, and the climate model used was not appropriate (Zenghelis, 2010). Another argument was that a CBA is unable to incorporate and measure the most important benefits of climate change mitigation (Ackerman, 2008). Similarly, there was a debate between a group of scientists associated with the Club of Rome, an interdisciplinary think tank, who were authors of the "Limits to Growth" (Meadows, Meadows, Randers, & Behrens, 1972), and corresponding updates about how to assess growth (Beinecke et al., 2012). For non-academics, the corresponding debates sound strange. However, the controversies make it clear that, in particular in the case of (high levels of) uncertainty, there is more than just academic evidence to be considered. It is not easy to assess the impact of effects over time. Time frames have the potential to change a programme's impact. Conclusions can be rather different for stakeholders with different time preferences. For decision-makers, a much more comprehensive picture of evidence is needed than just incomplete summary measures, such as those provided by the CC.

Third, it was also criticised that the aim of the CC is to break down independent silos of funding and priorities by simultaneously choosing between a variety of problems and solutions. The interrelatedness of projects and outcomes is not considered at all

(Greenwood, 2010). This problem has been intensively discussed in DC for many years. However, providers of DC are still focussed on that kind of thinking.

Fourth, CC rankings generally do not fit within the window of opportunity in decision-making processes. Usually, governments are incapable of redefining their budgets on a large scale. Public-spending decisions often take several years. Therefore, evidence provided by the CC is probably only useful for awareness-raising.

6 Potential of evidence in development cooperation policy

6.1 Interpretation and discussion of results

Based on the literature, the following statements about the provision and use of evidence in DC are derivable.

For several years, there has been a debate among academics and think tanks about the value of evidence, appropriate channels and tools about how research can have a greater impact in policy-making. However, at the level of policy-making, the existing evidence culture is different between countries.

In Germany, the discussion of using more evidence in DC is just beginning. Compared to countries such as the Netherlands and the United Kingdom, there is more common consensus among the political parties that there is a need for DC, and therefore less of a need to justify aid (Klasen, 2017). Nevertheless, not all lip service is implemented into actual policy-making.

6.1.1 Scope of evidence

The ongoing discussion at a higher level about including more evidence in DC is mostly based on two *types of evidence*: academic and practical-technical evidence. In this regard, two issues are targeted:

1. Improving knowledge transmission from available evidence into policy- and decision-making (see e.g. Ward, 2017). This discussion is part of a broader debate in the academic community. There is already much literature about identifying transfer barriers, both on the side of providers and users. There are many concepts available with suggestions about how to make possible improvements. These approaches are important, but a critical discussion is needed about whether or not a concrete concept is useful in a particular context. In addition, the context of settings is very dynamic and complex. However, awareness-raising and capacity-building are important first steps.

2. Enhancing the creation of evidence based on results management and results-oriented approaches. The performance of RBM among countries is rather diverse, according to reviews provided by the OECD. This is the consequence of corresponding target systems, which are complex to different degrees and lead to problems in identifying appropriate indicators and measurement problems. There is a lack of knowledge about results-based approaches. Many of them are still in the pilot stage.

The direct inclusion of social evidence and civil society's experiences, in particular, is dependent on the setting. For example, there is evidence that it is sometimes used in the health sector because people are understood to be important stakeholders. At the high level, it is considered by incorporating (representatives of) civil society in policy-related decision-making. At the operational level, there is more space for taking social evidence into account.

The practice of DC shows that the "call for more academic evidence" is often too narrow. Strong and intensive collaboration between the operating units of implementing organisations and academics and an open dialogue can provide a win-win situation for all partners and guarantee valid evidence, which is an important source for learning at the operational level. In this regard, the relationship between the providers and users of evidence should not be understood as a one-way process. Feedback loops, for example between providers and users of evidence, are very important. This is one of the basic approaches of project management. Information provides starting points for learning and for adjustments of activities. This idea is also an important backbone of results-based approaches.

The current debate about evidence is strongly biased in terms of academic evidence. For academics, there is a need for more research funding for providing this type of evidence. However, providing academic evidence is probably expensive. Research projects often have a limited time horizon. Therefore, it is important for transferring them into practical routines that might also be restricted by potential financial issues. For providing the best available evidence for decision-making, it is thus important to extend the available evidence space by getting access to all available types of evidence, including social evidence. The latter could be included by considering social media. This goes along with the interests of policy-makers, who prefer to utilise a variety of sources of knowledge for making their decisions (Cairney & Oliver, 2017).

6.1.2 Strength and quality of evidence

Over time, the *quality* of available academic and practical evidence has improved in a striking way. There are now many evaluations available that rest on rigorous methods. The current debate is focussed on quantitative evidence. In DC, also qualitative results are of high relevance. There is a danger that the concept of evidence of hierarchy is transferred without closer reflection to the area of DC. Because of quality issues, researchers have an interest in conducting quantitative studies with a high number of observations. This could be a problem in terms of small, disadvantaged groups for whom data are not collected. Therefore, academic evidence is not available. This is a great problem related to the "Leave No One Behind" initiative.

There are many differences between applying the concept of EBM and using evidence in DC: The former is aligned at the micro-level, covers only one discipline and studies are conducted with a more homogenous design. The latter is more located at the macro level, covers more disciplines and corresponding studies are particularly unstandardised because of more complex settings. However, the absolute quality level of evidence may not be mixed up with the perceived strength of quality that policy-makers perceive. Especially the provision of social evidence is often connected with particular interests. In this regard, the ways by which a stakeholder is provided, their power and the way this is done are all very important.

6.1.3 Evidence-orientation

In recent years, the use of evidence in DC has increased a lot. In comparison to the past, the level of usage has increased substantially. Currently, the orientation mostly corresponds to the level of being *evidence-supported*. In particular, many factors have influence on the usage:

- **The call for evidence by political decision-makers is sometimes motivated by the decision to justify decisions already taken**. There is some evidence that setting evidence on the agenda of policy-makers was the result of strategic thinking and not a move towards a common good. This issue is discussed as policy-based evidence (Strassheim & Kettunen, 2014).

- **Evidence is used more strongly in more advanced settings**. In partner countries, there is often a lower level of awareness of the usefulness of evidence. Frequently, the capacity related to the finding, interpretation and use of evidence is missing.

- **There is an abundance of high-quality evidence in some areas, but large evidence gaps in other areas**. Over time, much evidence has been collected in the fields of health and education, but other policy areas have been neglected. For example, there is only limited evidence on institution-building as a consequence of the non-applicability of RCTs in this area.

- **The closer that a decision-making body is located to the political space, the less willing it is to consider evidence analytically.** For them, evidence is recognised as a tool for better accountability and justification of promoted policies. However, the level of willingness might be increased by an overall push, as initiated by the OECD.

- **The basic allocation of funds from DC is based on political priorities and not on value-for-money considerations,** as suggest by the Copenhagen Consensus. At the downstream levels, evidence is taken into account to a greater extent. At the level of more concrete DC interventions, experiences gained are considered, for example in terms of used channels and implemented activities. At the level of multilateral DC, the level of confidence in the activities of the implementing bodies – including the Global Fund and GAVI – plays a crucial role. Those partners must account for the allocation of funds.

- **Competition among implementing bodies for funds increases evidence-orientation.** Organisations such as Germany's GIZ, which competes with other providers, have strong incentives to justify the success of their projects by means of results management and impact evaluations. By using high-quality methods, the perceptions of the soundness of their own activities can be improved. Evidence is used in terms of accountability and for supporting further applications. However, in the context of German DC, the corresponding transparency is limited. Providers do not have strong incentives to act in a transparent way, and there is also no obligation to publish all results. Currently, there are incentives not to publish "too much". For example, project proposals and reports are often not published; GIZ only publishes summaries of evaluations, and people interested in project-related information do not know how to find it on the homepage and cannot get it.

- **There is a tendency to hide unwanted evaluation results.** In the domain of research, it is often not possible to publish the results of studies that do not report significant

impacts. Similarly, project leaders worry about obtaining approval for future projects. Sometime policy-makers do not have an interest in publishing critical studies. This could be a big problems for research institutes that are dependent on policy-makers. These issues can have adverse effects on learning from experience.

- **Increasingly, NGOs see the importance of using available evidence in their demands**. In the age of the internet and new types of ICT channels, moral appeals for providing aid are not enough. Because a large number of NGOs are competing for scarce resources, selective evidence that mostly confirms their own views is used in their campaigns for proving the corresponding needs. In this regard, credible information and the reputation of actors is very important (Green, 2017; Guijt & de Goede, 2017).

6.2 Future potential

Similar to areas of activity such as health and education, the importance of evidence in DC will increase in the future. In order to improve the relevance of evidence in policy-making, the existing creation and provision of evidence must be strengthened. Types and forms of evidence transmission for practical use – including joint production with policy-makers – have to be improved. Policy-makers must also develop a better understanding of the usefulness of evidence, at least for reasons of accountability and better performance of activities targeting their political goals.

Evidence that is relevant in DC does not just refer to academic evidence, but academics do play an important function in this policy area. Currently, researchers often do not have strong incentives to write policy-related papers. The current system, the framework and the incentives for academics for managing and organising their work is based on impact points. As a matter of anticipated rewards, researchers are often more motivated to publish theoretical papers with only limited practical value. Even though this system is dissociated to a high degree from public regulation, rewards and incentives for providing more practice-oriented output can be implemented. Strengthening institution-building is very important to improve the use of evidence (Parkhurst, 2016).

The case regarding the insufficient exposure of World Bank policy papers has shown that policy advice based on a linear model between researchers who provide evidence and policy-makers as customers has shortcomings. In order to increase the awareness of academic results, policy-related academic results must be provided to policy-makers in a reasonable way. Policy-makers do not read academic papers and do not get excited by three stars highlighting statistical significance in a regression table. They want to be informed about the corresponding impact of interventions. Nowadays, policy briefs are often not seen by the corresponding target group. Results delivered by social media that are accompanied by more explicit engagement and communication strategies are needed. In this regard, credibility is central in getting policy-makers to use research findings (Pittore, te Lintelo, Georgalakis, & Mikindo, 2017). Because policy-makers are an extremely heterogeneous group, the evidence provided must be more tailored. Moreover, the understanding of research and policy advice needs adjustment – towards the concept of transformative research with a transdisciplinary focus, including a close exchange with decision-makers.

Social actors must be integrated into the research process through the co-designing and co-production of knowledge.[7]

The huge increase in the number of impact evaluations in recent years has not been enough. The main challenge for impact evaluations is to produce still more studies. Global policy should not be based on a single study from a single country, but rather on a large number of studies confirming whether an intervention works or not, and how that impact varies according to context. The corresponding conclusions are normally only relevant for a special setting and cannot be generalised. The call for more RCTs also has its limits. They are not a panacea. In some settings, such as climate change, RCTs are not effective (Leigh, 2010). In general, the expected value of additional studies should be critically compared to their costs: There is a need to consider the economics of evidence.

In order to overcome the existing shortcomings of synthesising evidence, common measures must be standardised so that knowledge that is more valuable can be created. In addition, standard variables should be included to improve the comparison of study results. Based on that, it is possible to disseminate joint evidence.

At the level of ministries, administrative bodies and implementing organisations, ways to overcome current problems of RBM must be identified. The exchange of experiences on platforms provided by the OECD is valuable. However, as the example from the United Kingdom has proven, the implementation of an evidence culture might be helpful. To do this, there is a need for facilitators.

Policy-makers often have different views about the value of evidence – compared to academics. Because they balance various interests and a variety of positions rather than focussing on details from a disciplinary perspective, they frequently have broader views. Offering policy-makers a more comprehensive evidence picture than just special results based on disciplinary tunnel vision might be a first important step for acknowledging the value of evidence.

For assessing the potential of evidence in DC, the sustainability of projects and programmes needs to be considered. The BMZ-funded programme RBF4MNH was rated a success, (White-Kaba, 2017) which was confirmed by accompanying research (Brenner & De Allegri, 2016). All stakeholders acknowledged the power of evidence and affirmed their interest in supporting that initiative. Therefore, BMZ prolonged programme funding; the government included the programme into the national health-sector strategy and asked for some donor basket funding until the question of future funding could be clarified. However, Malawi is heavily dependent on external financing; the question is whether there will be enough resources for running that programme in the future. In addition, the revealed positive evidence is dependent on the setting and the incentive structure that was given in the past. However, these configurations can change, and thus too the overall positive results. There is also the danger that RBF4MNH is understood as a lighthouse project and funding is concentrated on that initiative, meaning that money is not available for alternative projects. In a nutshell, there are many open questions about sustainability.

7 This is the case in the concept of scientific policy advice of the German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).

To sum up, the potential of evidence in DC depends strongly on many variables. To increase it in a substantial way, establishing a sound evidence-oriented culture in policy-making settings is a sine qua non. To achieve this, basing policy advice on transformative research, integrating all societal actors in the process of research by co-design and co-production of knowledge are extremely helpful.

6.3 Strength and weaknesses

The discussion about using evidence in DC is relatively new. Compared to the existing literature, this paper offers a more comprehensive look at the provision and use of evidence based on a conceptual framework. However, there are also some limitations in this paper. In general, "evidence" – as used in the literature – is a rather fuzzy term. There is a different understanding, and sometimes statements about the use of evidence is rather rhetorical. Based on a literature review, a full assessment based on the conceptual framework was not possible. Therefore, an empirical study is necessary in which DC stakeholders answer questions about their perceptions and the concrete use of evidence – compared to other issues that are important in decision-making.

7 Conclusions and recommendations

It has been the main aim of this paper to analyse the provision and use of evidence in DC. In particular, its potential in the near future is considered in terms of how – and in what context – it can be used. Evidence is a term with several meanings and connotations. The term is often used interchangeably for empirical data, information, statistics, metrics and measurements, but often evidence expresses the ability to draw conclusions based on available data, information and/or knowledge that an activity works. To bridge the provision and use of evidence, a model consisting of three dimensions was introduced: first, the scope of using different types of evidence; second, the strength and quality of evidence; third, the relevance of evidence in policy-making.

Some 10 years ago, the quality of available evidence about what works in DC was poor. Activities and interventions in the field of DC were input-dominated and driven by concepts and thoughts about how development should work. In particular, there was a lack of performance- and results-based data, which provide the floor for establishing evidence-oriented information. However, evidence was available and also used – but more on an anecdotal, narrative level.

Since then, the provision of evidence in DC has improved. Management practices and procedures with an explicit focus on results and results-oriented aid approaches have been started. An evaluation culture was also established in the field of DC in pioneering countries. The corresponding number and quality of studies providing evidence has increased. International networks and organisations now offer capacity-building for providing more and better evidence. There are institutes and persons who aim at increasing awareness for using more and better evidence at the level of decision-making and administrative bodies. There are training programmes for developing an understanding about how to get evidence

and how to use it. However, the adoption of evidence in policy-making has been mixed. Mostly, evidence matters because policy-making is supported by it.

There are differences between countries, the levels of policy-making, and the willingness to learn and implement evidence. In addition, the political economy must be considered. In spite of international agreements about more aid effectiveness and lip service about the "importance of more evidence-based policy-making", political actors often are not interested in taking the whole menu of provided evidence into account. Ranking schemes – including the value-for-money considerations of different interventions, as provided by the Copenhagen Consensus Centre – are welcome. However, such menus are often used only for reasons of cherry picking. Existing uncertainty about the future and serious doubts about the underlying calculations provided by scientists and think tanks can easily be used as an excuse for not considering the provided evidence at all.

Sure, more evidence-orientation is better. Nevertheless, it is an illusion to think about full EBPM, including the complete range of evidence components and a maximum quality level. From an economic perspective, maximising scope, strength and quality as well as evidence-orientation is also not reasonable because of the corresponding opportunity costs. Identifying sound evidence requires time, and sometimes quick responses are needed, for example in the case of humanitarian aid. However, evidence is needed – even in this case because, analogous to health care, the knowledge of "what works" can help save lives (de Geoffroy, Léon, & Beuret, 2015). The particular challenge is to provide the best available evidence in a timely manner, even in complex emergencies (Bradt, 2009).

The Copenhagen Consensus case study also illustrates that academic evidence is in some cases widely discussed. Recommendations derived from various academic disciplines are in some instances rather different and clash. Beyond that, based on their disciplinary backgrounds, academics have a different understanding about what is acceptable evidence for political action (Jasanoff, 1995). There are also social scientists who question research methods looking for clarity and precision. Radical social scientists such as Law (2004) argue that methods not only describe social realities but also help to create them. There is also a hinterland of realities – of manifest absences and otherness, resonance and patterns of one kind or another – already being enacted. Those cannot be ignored in the process of providing evidence.

Sometimes academic results can also be in conflict with certain value judgements. Research questions and strategies in disciplines are also based on certain world views with a certain Zeitgeist. It is possible that efficiency and effectiveness issues are in conflict with distributional questions.

In light of the above, there is a need to exploit the maximum amount of the available potential. On the supply side, structures for strengthening the provision of evidence must be improved. Therefore, helping factors need to be identified and targeted. This includes encouraging academics to focus their work more on policy issues and to improve the transferability of their results. Certainly, many more evaluation studies should be conducted using international standards. However, more is not always better. Because of high evaluation costs, the implementation can be limited to settings in which evidence is missing. It is important to share the results globally and provide open access. Small-scale studies should be avoided. The same principles are also valid in terms of monitoring and evidence

from results-oriented approaches. However, there is not only a need for more and better academic evidence, but also for collecting evidence from other sources, for example in terms of practical experience and civil society-based know-how.

Currently, much available evidence is not used for policy-making. At the level of potential users, it is important to create a more evidence-oriented culture. There is a need for more facilitators to translate academic evidence into the language of policy-makers and administrative bodies. In the past, many models for knowledge transmission had been developed (Newman et al., 2014). Against the background of concrete settings, appropriate concepts for pathways to success need to be identified – for example, direct communication with all stakeholders during all project stages – for best identifying relevant evidence that might be useful.

The message of this paper is NOT that every piece of research should be directed to a platform for optimal exposure and used by policy-makers. Researchers should NOT only anticipate what is written on the political agenda. Of course, topics like combating diseases such as HIV/AIDS, malaria and tuberculosis are of high relevance. However, focussing on such "high priority" topics has had some severe side effects, namely not addressing neglected diseases appropriately. To recognise future challenges, important topics and relevant action fields, basic and exploratory research is needed – more than ever. This is also the case in international development cooperation. However, it is important that policy-oriented studies be designed in a better way; identified barriers on the supply and demand sides should be better targeted so that the impact of research is improved. In the case of development cooperation, we are on a good path, but there is much more left to do.

Policy recommendations and lessons learnt

According to the results of the above analysis, the following recommendations for German DC are given:

1. **To increase the impact of German DC, a strong focus on evidence is needed.** Based on the particular issues of problem-setting, including the timescale, the types of evidence needed must be chosen. Therefore, it is helpful to establish a strong evidence- and results-based culture in all parts of the German DC system. The objective should be to incorporate evidence into decisions in a timely and efficient way.

2. **Striving to identify "what works" is central to the mission of German DC, and to those who are supported.** Thus, the use of evidence should be enshrined in all strategic papers of German DC.

3. **Specify the type of evidence used, as well its strengths, when making or proposing decisions.** The available evidence differs. Reflect upon and define the proper criteria for assessing the evidence's strengths, such as validity, timelines and reliability. This will help to define what the best available evidence is, according to each specific context and time frame.

4. **For supporting evidence-oriented policy-making, specific studies can be helpful.** Available evidence must be the backbone of policy-making. Supporting the bodies of

the German DC system in identifying the appropriate sources of evidence and initiating research in cases where there are existing evidence gaps in studies are also helpful.

5. **Strengthening German results management is important for accountability and improving decision-making.** A full-functioning results management that is capable of integrating all relevant results information is needed for sound and consistent DC policy, planning, budgeting and communication. A strong results system enables a better understanding of what drives outcomes and impacts in the context of broader development processes. Therefore, implementing a results framework based on the model of other donors could be an important contribution.

6. **Decision-makers in partner countries must be supported in using evidence in policy-making.** There is insufficient information about the impact of German aid in partner countries because partner countries are often not interested in evaluation. Achieving development goals is not only a question of money, but also of expertise and capacity. Therefore, access to evidence is necessary. However, there are some high barriers, including problems in understanding the complexity of evidence and a lack of analytical capacity. DC must address capacity gaps at the programme level. For example, monitoring and evaluation is highly complex if quantitative data must be analysed. The necessary statistical knowledge and software is often not available.

7. **To improve the impact in neglected policy areas, existing evidence gaps must be identified and addressed.** The SDGs are strongly interconnected. To design appropriate projects, there is a need for a better understanding of the effects of concrete interventions. For example, forest protection is among the most effective approaches we have to mitigate climate change. At the same time, agricultural land and forests provide food, livelihoods and fuel for billions of people. There are concerns that large-scale forest protection programming will have negative effects on food security and other aspects for people in low-income countries. In order to design interventions without adverse side effects, a collaboration between implementing bodies of German DC and research is needed.

8. **Increase the use of robust impact evaluations.** In German DC, impact evaluations are still rather rarely used. Currently, the collaboration between researchers and operative departments of implementing organisations is selective and mostly based on individual initiatives. There is still scepticism among some project leaders about the value of impact analysis. However, there is a broad consensus in development policy that the impact of interventions should be assessed using rigorous impact evaluations. Therefore, including more accompanying research is a reasonable objective. The motivation for conducting impact analyses could be improved if a special fund were created at BMZ or DEval. Project leaders could apply, and suppliers of evaluations can be selected though tendering. Based on own impact analyses, implementing organisations and DEval should also work on more systematic reviews.

9. **Extend the volume of research funding in the context of German DC.** More targeted funding of research implies more available evidence and greater impact of German DC.

10. **Conducting impact evaluations is no end in itself.** A much broader discussion about the results is necessary, so learning is possible. The value of evidence should be

emphasised through public campaigns initiated by institutions. More resources and time for dissemination strategies should be included in the case of research programmes.

11. **There is a need for more systematic learning.** Experiences from impact evaluation studies are currently provided in a selected way. To increase the impact, providers must improve their networking and pooling of available results. In this regard, improving institutional structures for promoting systematic learning is important.

12. **Include evidence as an explicit criterion in project selection.** Currently, project proposals are assessed in terms of their economic, socio-economic, socio-cultural, gender-specific and ecological feasibility. The recently implemented procedural reform was an improvement in this context. However, there is still a need to include evidence-based information more directly. In addition, there remains a lack of transparency. Practices implemented through other important OECD/DAC member countries can be used as a source for improving current procedures.

13. **Consider the use of evidence in decision-making as an institutional issue.** This could lead to the formulation of internal policy on guidance for addressing the multiple repercussions of such a commitment at different decision-making levels within an organisation. A well-designed knowledge management is an important basis for institutional learning.

References

Achinstein, P. (2001). *The book of evidence*. Oxford: Oxford University Press.

Ackerman, F. (2008). *Hot, it's not: Reflections on cool it, by Bjorn Lomborg*. Berlin: Springer.

Baker, M., & Salib, M. (2017). Cracking the evidence conundrum: Four ideas to get people to use evidence. *Learning Lab*. Retrieved from https://usaidlearninglab.org/lab-notes/cracking-evidence-conundrum-four-ideas-get-people-use-evidence

Bamberger, M. (2015). Management of complexity-responsive evaluations. In M. Bamberger, J. Vaessen, & E. Raimondo (Eds.), *Dealing with complexity in development evaluation: A practical approach* (pp. 48-59). London, New Dehli, Singapore: Sage Publications.

Bamberger, M., Vaessen, J., & Raimondo, E. (2015). *Dealing with complexity in development evaluation: A practical approach*. London, New Dehli, Singapore: Sage Publications.

Banerjee, A. V., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. New York, NY: Public Affairs.

Banks, G. (2009). *Evidence-based policy making: What is it? How do we get it?* Melbourne: Australian Government, Productivity Commission.

Barr, J., Rinnert, D., Lloyd, R., Dunne, D., & Henttinen, A. (2016). *The value of evaluation: Tools for budgeting and valuing evaluations*. Retrieved from https://assets.publishing.service.gov.uk/media/57a312d8e5274a31e0001b5a/Value_of_Evaluation_Discussion_Paper_-__Final_Version_for_Publication_03082016_clean.pdf

Bédécarrats, F., Guérin, I., & Rouband, F. (2017). All that glitters is not gold. The political economy of randomized evaluations in development. *Development and Change.* http://dx.doi.org/10.1111/dech.12378

Beinecke, F., Meadows, D., Randers, J., Harte, J., Harte, M. E., & Lomborg, B. (2012). Is growth good? Resources, development, and the future of the planet. *Foreign Affairs, 91*(5), 163-175.

Bernal, P., Celhay, P., & Martinez, S. (2018). *Is results-based aid more effective than conventional aid? Evidence from the health sector in El Salvador*. Retrieved from https://publications.iadb.org/bitstream/handle/11319/8750/is-results-based-aid-more-effective-than-convention-aid.PDF?sequence=1&isAllowed=y

Binnendijk, A. (2000). *Results based management in the development cooperation agencies: A review of experience.* Background report, DAC Working Party on Aid Evaluation. Paris: Organisation for Economic Co-operation and Development.

Blume, J. (2012). *Results-based approaches in German financial cooperation*. Retrieved from https://www.kfw-entwicklungsbank.de/Download-Center/Themenpublikationen/Results-based-approaches-in-German-Financial-Cooperation.pdf

Bradt, D. A. (2009). *Evidence-based decision-making in humanitarian assistance* (HPN Network Paper 67). London: Humanitarian Practice Network, Overseas Development Institute.

Brenner, S., & De Allegri, M. (Eds.). (2016). *Final results of the RBF4MNH impact evaluation.* Prepared by the Research Team Members of the University of Heidelberg and College of Medicine. University of Heidelberg.

Buetow, S., & Kenealy, T. (2000). Evidence-based medicine: The need for a new definition. *Journal of Evaluation in Clinical Practice, 6*(2), 85-92.

Cairney, P., & Oliver, K. (2017). Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? *Health Research Policy and Systems, 15*(1), 35.

Cameron, D. B., Mishra, A., & Brown, A. N. (2016). The growth of impact evaluation for international development: How much have we learned? *Journal of Development Effectiveness, 8*(1), 1-21.

Camfield, L., Duvendack, M., & Palmer-Jones, R. (2014). Things you wanted to know about bias in evaluations but never dared to think. *IDS Bulletin, 45*(6), 49-64.

Cohen, C., & Werker, E. D. (2008). The political economy of "natural" disasters. *Journal of Conflict Resolution, 52*(6), 795-819.

Carden, F. (2009). *Knowledge to policy: Making the most of development research*. Ottawa: International Development Research Centre.

Clark, W. C., van Kerkhoff, L., Lebel, L., & Gallopin, G. C. (2016). Crafting usable knowledge for sustainable development. *Proceedings of the National Academy of Sciences, 113*(17), 4570-4578.

Copenhagen Consensus Centre. (2016). *Bangladesh priorities eminent panel findings*. Retrieved from http://www.copenhagenconsensus.com/sites/default/files/bangladesh_priorities_outcome_document_final.pdf

Davies, H. T. O., Nutley, S. M., & Smith, P. C. (Eds.). (2000). *What works? Evidence-based policy and practice in public services*. Bristol, UK: Blackwell Publ. Ltd.

De Geoffroy, V., Léon, V., & Beuret, A. (2015). *Evidence-based decision making for funding allocations*. Retrieved from https://www.ghdinitiative.org/assets/files/Activities/Our%20Work/Evidence_Based_Study_Final_15_10_01.pdf

Department for International Development. (2011). *UK aid: Changing lives, delivering results*. London: Author.

Deutsche Gesellschaft für internationale Zusammenarbeit. (2013). *Monitoring an evaluation unit – in focus: Knowing what works*. Eschborn: Author.

Dicks, L. V., Walsh, J. C., & Sutherland, W. J. (2014). Organising evidence for environmental management decisions: A "4S" hierarchy. *Trends in ecology & evolution, 29*(11), 607-613.

Dobrow, M. J., Goel, V., & Upshur, R. (2004). Evidence-based health policy: Context and utilisation. *Social science & medicine, 58*(1), 207-217.

Doemeland, D., & Trevino, J. (2014). *Which World Bank reports are widely read?* Retrieved from http://documents.worldbank.org/curated/en/387501468322733597/pdf/WPS6851.pdf

Donaldson, S. I., Christie, C. A., & Mark, M. M. (Eds.). (2009). *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage.

Evidence-Based Medicine Working Group. (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *Jama, 268*(17), 2420-2425.

Eyben, R. (2015). Uncovering the politics of evidence and results. In R. Eyben, I. Guijt, C. Roche, & C. Shutt (Eds.), *The politics of evidence and results in international development: Playing the game to change the rules* (pp. 19-38). Rugby, UK: Practical Action Publishing.

Faust, J. (2017, November 30). *(Quasi-)experimental impact evaluation in the German development cooperation: Usefulness, potential of big data and future cooperation*. DEval workshop. Bonn: German Institute for Development Evaluation.

Fourie, W. (2017). *Six barriers that make it difficult for African states to use research for policy*. Retrieved from https://theconversation.com/six-barriers-that-make-it-difficult-for-african-states-to-use-research-for-policy-86492

Gaarder, M. (2014). *Beyond good intentions: Designing an evidence-based approach to aid*. Retrieved from http://ieg.worldbankgroup.org/blog/beyond-good-intentions-designing-evidence-based-approach-aid

German Federal Ministry for Economic Co-operation and Development. (2008). *Leitlinienfür die bilaterale Finanzielle und technische Zusammenarbeit mit Kooperationspartnern der deutschen Entwicklungszusammenarbeit* (Vol. 165). Bonn: Author.

Global Partnership for Effective Development Co-operation. (2016). *Making development co-operation more effective: 2016 progress report*. Retrieved from http://www.undp.org/content/dam/undp/library/development-impact/--2016%20progress%20report--Final%20(e-book).pdf

Green, D. (2017). *How can NGOs get better at using evidence to influence governments and companies?* Retrieved from http://oxfamblogs.org/fp2p/how-can-ngos-get-better-at-using-evidence-to-influence-governments-and-companies/

Greenwood, D. T. (2010). Solutions for the world's biggest problems: Costs and benefits. *Eastern Economic Journal, 36*(4), 539-541.

Grittner, A. M. (2013). *Results-based financing: Evidence from performance-based financing in the health sector* (Discussion Paper 6/2013). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).

Guijt, I., & de Goede, I. (2017). Evidence for influencing: Why we're excited about the first Oxfam Research Network conference. Retrieved from https://views-voices.oxfam.org.uk/methodology/research/2017/10/evidence-influencing-excited-first-oxfam-research-network-conference/

Guijt, I., & Roche, C. (2014). Does impact evaluation in development matter? Well, it depends what it's for! *The European Journal of Development Research, 26*(1), 46-54.

Guo, J., Hepburn, C. J., Tol, R. S., & Anthoff, D. (2006). Discounting and the social cost of carbon: A closer look at uncertainty. *Environmental Science & Policy, 9*(3), 205-216.

Hamaide, B., & Boland, J. J. (2006). Cost-benefit analysis is useful in evaluating climate change policies. *World Resource Review, 18*(2), 279-290.

Hewman, K., Capillo, A., Famurewa, A., Nath, C., & Siyanbola, W. (2013). *What is the evidence on evidence-informed policy making? Lessons from the International Conference on Evidence-Informed Policy Making.* Retrieved from http://www.inasp.info/uploads/filer_public/2013/04/22/what_is_the_evidence_on_eipm.pdf

Hjørland, B. (2011). Evidence-based practice: An analysis based on the philosophy of science. *Journal of the Association for Information Science and Technology, 62*(7), 1301-1310.

Holzapfel, S. (2014). *The role of indicators in development cooperation: An overview study with a special focus on the use of key and standard indicators* (Studies 81). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).

Holzapfel, S., & Janus, H. (2015). *Improving education outcomes by linking payments to results* (Discussion Paper 2/2015). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).

Howlett, M., & Craft, J. (2013). Policy advisory systems and evidence-based policy: The location and content of evidentiary policy advice. In S. P. Young (Ed.), *Evidence-based policy-making in Canada: A multidisciplinary look at how evidence and knowledge shape Canadian public policy* (pp. 27-44). Don Mills, Ontario: Oxford University Press.

Janus, H., & Holzapfel, S. (2016). *Results-based approaches in agriculture: What is the potential?* (Discussion Paper 25/2016). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).

Jasanoff, S. (1987). Contested boundaries in policy-relevant science. *Social Studies of Science, 17*(2), 195-230.

Jasanoff, S. (1995). Product, process, or programme: Three cultures and the regulation of biotechnology. In M. Bauer (Ed.), *Resistance to new technology: Nuclear power, information technology and biotechnology* (pp. 311-331). Cambridge: Cambridge University Press.

Jasanoff, S. (2005). Restoring reason: Causal narratives and political culture. In B. Hutter & M. Power (Eds.), *Organizational encounters with risk* (pp. 209-232). Cambridge: Cambridge University Press.

Jasanoff, S. (2006). Just evidence: The limits of science in the legal process. *The Journal of Law, Medicine & Ethics, 34*(2), 328-341.

Kapiriri, L., Sinding, C., & Arnold, E. (2017). How do development assistance partners conceptualise and prioritise evidence in Priority Setting (PS) for health programmes relevant to low-income countries? A qualitative study. *Evidence & Policy: A Journal of Research, Debate and Practice, 13*(2), 255-273.

Keijzer, N., & Klingebiel, S. (2017). *Realising the data revolution for sustainable development: Towards capacity development 4.0*. Retrieved from https://www.paris21.org/sites/default/files/CapacityDevelopment4.0_FINAL_0.pdf

Kemper, N., & Sieler, S. (2016). *Results-based approaches in international development cooperation*. Retrieved from https://www.kfw-entwicklungsbank.de/PDF/Download-Center/PDF-Dokumente-Development-Research/2016-11-17_EK_EBA_EN.pdf

KfW Development Bank. (2016). *Ex post evaluation Zimbabwe*. Retrieved from https://www.kfw-entwicklungsbank.de/PDF/Evaluierung/Ergebnisse-und-Publikationen/PDF-Dokumente-R-Z_EN/Simbabwe_OVC_2015_E.pdf

KfW Development Bank. (2017). *14th evaluation report 2015-2016*. Frankfurt am Main: Author.

Klasen, S. (2017, November 30). *(Quasi-)experimental impact evaluation in the German development cooperation: Usefulness, potential of big data and future cooperation*. DEval workshop. Bonn: German Institute for Development Evaluation.

Klingebiel, S. (2012). *Results-based aid (RBA): New aid approaches, limitations and the application to promote good governance* (Discussion Paper 14/2012). Bonn: German Development Institute / Deutsches Institut für Entwicklungspolitik (DIE).

Klingebiel, S. (2014). *Development cooperation: Challenges of the new aid architecture*. Basingstoke: Palgrave Macmillan.

Klingebiel, S., & Janus, H. (2014). Results-based aid: Potential and limits of an innovative modality in development cooperation. *International Development Policy, 6*(1).

Klingebiel, S., Mahn, T., & Negre, M. (Eds.). (2016). *The fragmentation of aid: Concepts, measurements and implications for development cooperation.* Basingstoke: Palgrave Macmillan.

Krisch, F., Schmitt, J., & Dörr, U. (2015). *Begleitende Maßnahmen der allgemeinen Budgethilfe in Subsahara-Afrika*. Retrieved from http://www.deval.org/files/content/Dateien/Evaluierung/Berichte/DEval_Begleitende_Massnahmen_Budgethilfe_Gesamtbericht.pdf

Lavis, J. N., Ross, S. E., & Hurley, J. E. (2002). Examining the role of health services research in public policymaking. *The Milbank Quarterly, 80*(1), 125-154.

Law, J. (2004). *After method: Mess in social science research*. London: Routledge.

Leigh, A. (2010). Evidence-based policy: Summon the randomistas? In Productivity Commission (Ed.), *Strengthening evidence-based policy in the Australian Federation, Volume 1: Proceedings, roundtable proceedings* (pp. 215-226). Canberra: Productivity Commission.

Levine, R., & Savedoff, W. D. (2016). *Impact evaluation: How the wonkiest subject in the world got traction*. Retrieved from http://blogs.3ieimpact.org/impact-evaluation-how-the-wonkiest-subject-in-the-world-got-traction/

Lomborg, B. (2004a). Need for economists to set global priorities. *Nature, 431*(7004), 17.

Lomborg, B. (Ed.). (2004b). *Global crisis, global solutions*. Copenhagen: Copenhagen Consensus Center.

Lomborg, B. (2007a). *Cool it: The skeptical environmentalist's guide to global warming*. New York, NY: Vintage Books.

Lomborg, B. (2007b). *Solutions for the world's biggest problems: Costs and benefits*. Cambridge: Cambridge University Press.

Lomborg, B. (Ed.). (2010). *Smart solutions to climate change: Comparing costs and benefits*. Cambridge: Cambridge University Press.

Lomborg, B. (2012). *Rethink HIV: Smarter ways to invest in ending HIV in sub-Saharan Africa*. Cambridge: Cambridge University Press.

Lomborg, B. (2015). *The Nobel laureates' guide to the smartest targets for the world: 2016-2030*. Copenhagen: Copenhagen Consensus Center.

Lomborg, B. (2017). Making government smarter: How to set national priorities. *Foreign Affairs, 96*, 90-98.

Lundahl, M. (2013). *The political economy of disaster: Destitution, plunder and earthquake in Haiti* (Vol. 61). Oxford: Routledge.

Mallett, R., Hagen-Zanker, J., Slater, R., & Duvendack, M. (2012). The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness, 4*(3), 445-455.

Marmot, M. G. (2004). Evidence-based policy or policy-based evidence?: Willingness to take action influences the view of the evidence – look at alcohol. *BMJ: British Medical Journal, 328*(7445), 906.

McGillivray, M., Feeny, S., Hermes, N., & Lensink, R. (2005). *It works; it doesn't; it can, but that depends...: 50 years of controversy over the macroeconomic impact of development aid* (9291907391). Retrieved from https://www.wider.unu.edu/sites/default/files/rp2005-54.pdf

McNulty, J. (2012). Symbolic uses of evaluation in the international aid sector: Arguments for critical reflection. *Evidence & Policy: A Journal of Research, Debate and Practice, 8*(4), 495-509.

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W. (1972). *The limits to growth*. New York, NY: Universe Books.

Miles, B., & Morse, S. (2007). The role of news media in natural disaster risk and recovery. *Ecological Economics, 63*(2), 365-373.

Monaghan, M. (2013). Sources and resources: An interview with ... Dr Howard White by Mark Monaghan. *Evidence & Policy, 9*(1), 131-136.

Musgrove, P. (2011). *Rewards for good performance or results: A short glossary*. Washington, DC: World Bank.

Newman, K., Evans, W., Dercon, S., Whitty, C., Harvey, M., Vince, J., & Rimmer, J. (2014). *What is the evidence on the impact of research on international development?* London: Department for International Development.

Nutley, S., & Webb, J. (2000). Evidence and the policy process. In H. T. O. Davies, S. M. Nutley, & P. C. Smith (Eds.), *What works? Evidence-based policy and practice in public services* (pp. 13-41). Bristol, UK: Blackwell Publ. Ltd.

OCEBM Levels of Evidence Working Group. (2011). The Oxford 2011 levels of evidence. Retrieved from http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf

OECD (Organisation for Economic Co-operation and Development). (2006). *Peer review: Germany*. Paris: Author.

OECD. (2014a). *Measuring and managing results in development co-operation: A review of challenges and practices among DAC members and observers*. Paris: Author.

OECD. (2014b). *OECD development co-operation peer reviews: United Kingdom*. Paris: OECD Publishing.

OECD. (2015). *OECD development co-operation peer reviews: Germany*. Paris: OECD Publishing.

OECD. (2016a). *Evaluation systems in development co-operation: 2016 review*. Paris: OECD Publishing.

OECD. (2016b). *OECD DAC Peer review USA*. Paris: OECD Publishing.

OECD. (2017). *Development co-operation report 2017: Data for development*. Paris: Author.

OECD. (s.a.). *Results-based decision making*. Retrieved from http://www.oecd.org/development/results-measurement-and-management.htm

OECD/DAC (Development Assistance Committee). (2002). *Glossary of key terms in evaluation and results-based management*. Working Party on Aid Evaluation. Paris: Author.

Oliver, K., Innvær, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC health services research, 14*(1), 2.

Oliver, K., Lorenc, T., & Innvær, S. (2014). New directions in evidence-based policy research: A critical analysis of the literature. *Health Research Policy and Systems, 12*(1), 34.

Parkhurst, J. (2016). *The politics of evidence: From evidence-based policy to the good governance of evidence*. Oxford: Routledge.

Paul, E., Albert, L., Bisala, B. N. S., Bodson, O., Bonnet, E., Bossyns, P., . . . Eclou, D. S. (2018). Performance-based financing in low-income and middle-income countries: Isn't it time for a rethink? *BMJ Global Health, 3*(1), e000664.

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. Thousand Oaks, CA: Sage.

Perakis, R., & Savedoff, W. (2015*). Does results-based aid change anything? Pecuniary interests, attention, accountability and discretion in four case studies* (CGD Policy Paper 52). Washington, DC: Center for Global Development.

Pittore, K., te Lintelo, D. J., Georgalakis, J., & Mikindo, T. (2017). Choosing between research rigour or support for advocacy movements, a false dichotomy? *IDS Bulletin, 47*(6). Brighton, UK: Institute of Development Studies.

Rinnert, D., & Brower, L. (2017). *How is evidence actually used in policy-making? A new framework from a global DFID programme*. Retrieved from https://blogs.oxfam.org/en/content/how-evidence-actually-used-policy-making-new-framework-global-dfid-programme

Roberts, D., & Khattri, N. (2012). *Designing a results framework for achieving results: A how-to guide*. Washington, DC: International Bank for Reconstruction and Development (World Bank).

Rogers, P., Hawkins, A., McDonald, B., Macfarlan, A., & Milne, C. (2015). *Choosing appropriate designs and methods for impact evaluation*. Retrieved from https://industry.gov.au/Office-of-the-Chief-Economist/Publications/Documents/Impact-evaluation-report.pdf

Rudolph, A., von Schiller, A., & Strupat, C. (2017). *Lernen aus Evidenz – Wie Wirkungsanalysen besser genutzt werden können* (Die aktuelle Kolumne). Retrieved from https://www.die-gdi.de/die-aktuelle-kolumne/article/lernen-aus-evidenz-wie-wirkungsanalysen-besser-genutzt-werden-koennen/

Sachs, J. D. (2004). Solving global crises: Economists alone are not enough. *The Lancet, 364*(9451), 2087-2088.

Savedoff, W. D., Levine, R., & Birdsall, N. (2006). *When will we ever learn? Improving lives through impact evaluation*. Washington, DC: Center for Global Development.

Scott-Findlay, S., & Pollock, C. (2004). Evidence, research, knowledge: A call for conceptual clarity. *Worldviews on Evidence-Based Nursing, 1*(2), 92-97.

Segone, M. (2007). Why evaluate? The evolution of the evaluation function. In M. Segone (Ed.), *New trends in development evaluation* (pp. 8-30). New York, NY: UNICEF Regional Office for CEE/CIS and the International Program Evaluation Network.

Segone, M., & Ocampo, A. (Eds.). (2006). *Creating and developing evaluation organizations, lessons learned from Africa, Americas, Asia, Australasia and Europe*. Lima: desco.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations. Report of a study commissioned by the Department for International Development*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf

Stewart, R. (2015). A theory of change for capacity-building for the use of research evidence by decision makers in southern Africa. *Evidence & Policy: A Journal of Research, Debate and Practice, 11*(4), 547-557.

Stewart, E., & Smith, K. E. (2015). "Black magic" and "gold dust": The epistemic and political uses of evidence tools in public health policy making. *Evidence & Policy: A Journal of Research, Debate and Practice, 11*(3), 415-437.

Strassheim, H., & Kettunen, P. (2014). When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. *Evidence & Policy: A Journal of Research, Debate and Practice, 10*(2), 259-277.

The Economist. (2016). *Developing Bangladesh: How to spend it*. Retrieved from https://www.economist.com/news/finance-and-economics/21698302-ambitious-attempt-work-out-best-use-scarce-resources-how-spend-it

Tytler, R. (2001). Dimensions of evidence, the public understanding of science and science education. *International Journal of Science Education, 23*(8), 815-832.

Uneke, C., Ezeoha, A., Ndukwe, C., Oyibo, P., Onwe, F., Igbinedion, E., & Chukwu, P. (2011). Individual and organisational capacity for evidence use in policy making in Nigeria: An exploratory study of the perceptions of Nigeria health policy makers. *Evidence & Policy: A Journal of Research, Debate and Practice, 7*(3), 251-276.

United Kingdom. (2014). *What works review of the use of evidence in the Department for International Development.* Evidence into Action and What Works Team. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/340971/What_Works_Review_DFID.pdf

United Nations Development Programme. (2007). *MDG-based national development strategies*. Poverty Group Bureau for Development Policy. New York, NY: Author.

United States Agency for International Development. (2017). *ADS chapter 201: Program cycle operational policy*. Retrieved from https://www.usaid.gov/sites/default/files/documents/1870/201.pdf

United States Government Accountability Office. (2017). *2017 survey of organizational performance and management issues*. Retrieved from https://www.gao.gov/special.pubs/gao-17-776sp/resultsall.htm

Vaessen, J., Raimondo, E., & Bamberger, M. (2015). Impact evaluation approaches and complexity. In M. Bamberger, J. Vaessen, & E. Raimondo (Eds.), *Dealing with complexity in development evaluation: A practical approach* (pp. 62-87). London, New Dehli, Singapore: Sage Publications.

Vähämäki, J., Schmidt, M., & Molander, J. (2011). *Results-based management in development cooperation.* Stockholm: Riksbankens Jubileumsfond.

Waddington, H., White, H., Snilstveit, B., Hombrados, J. G., Vojtkova, M., Davies, P., . . . Petticrew, M. (2012). How to do a good systematic review of effects in international development: A tool kit. *Journal of Development Effectiveness, 4*(3), 359-387.

Ward, V. (2017). Why, whose, what and how? A framework for knowledge mobilisers. *Evidence & Policy: A Journal of Research, Debate and Practice, 13*(3), 477-497.

White, H. (2014). Current challenges in impact evaluation. *The European Journal of Development Research, 26*(1), 18-30.

White-Kaba, M. (2017). *How rewards improve health practice in Malawi: Learnings from a maternal and newborn Health Initiative*. Retrieved from http://health.bmz.de/ghpc/case-studies/how_rewards_improve_health_practice_malawi/RBF_Malawi_6.pdf

Witter, S., Kardan, A., Scott, M., Moore, L., & Shaxson, L. (2017). Generating demand for and use of evaluation evidence in government health ministries: Lessons from a pilot programme in Uganda and Zambia. *Health Research Policy and Systems, 15*(1), 86.

Wollmann, H. (2017). Entwicklungslinien von Evaluation und Evaluationsforschung in Deutschland. Zwischen Vorgeschichte, Aufbruch und Konsolidierung. *Zeitschrift für Evaluation, 16*(2), 33-53.

Young, J. (2006). Bridging research and policy: Insights from 50 case studies. *Evidence & Policy: A Journal of Research, Debate and Practice, 2*(4), 439-462.

Young, S. P. (2013). *Evidence-based policy-making in Canada: A multidisciplinary look at how evidence and knowledge shape Canadian public policy*. Don Mills, Ontario: Oxford University Press.

Young, K., Ashby, D., Boaz, A., & Grayson, L. (2002). Social science and the evidence-based policy movement. *Social Policy and Society, 1*(3), 215-224.

Zenghelis, D. (2010). Book review: Smart solutions to climate change: Comparing costs and benefits, edited by Bjørn Lomborg. *Journal of Environmental Investing, 1*(2), 79-89.

Appendix 1: Using evidence in different policy areas – some examples			
Policy area	**Approach**	**Strand and level**	**Explanation**
Health	Evidence-based medicine (EBM)	Strand 1: practice Micro-level	Focussed on the physician–patient relationship. The quality of available evidence as a reference for guiding health-related intervention is based on a hierarchy of (mostly quantitative) methods. Randomised control trials and their systematic reviews are understood as methods for proving best evidence. There is a particular emphasis on effectiveness. EBM aims at ensuring that it has a direct impact on practice by exhorting practitioners to replace bad interventions so as to improve outcomes. Based on evidence, intervention strategies at the macro-level can be identified.
	Evidence-based (public) health policy	Strand 2: policy-making Macro-level	Use of EBM, evidence from public health (if applicable) in combination with economic study results, e.g. health economic evaluation (cost-effectiveness analysis, cost-utility analysis). This approach aims to improve the allocation of funds in the health sector, including priority-setting.
Education	Evidence-based education policy	Strand 2: policy-making Macro-level	Covers all aspects of education – from classroom practice to policy-making. It is derived from empirical qualitative and quantitative research and signifies the idea that educational policy should be guided by the best evidence about the likely effects.
	Evidence-based education practice	Strand 1: practice Micro-level	Teaching is interpreted similar to health care; appropriate teaching concepts based on empirical insights are implemented in educational practice.
Public administration	Evidence-based budgeting	Strand 1: practice Micro-/meso-level	Performance budgeting links the funds allocated to measurable results, replacing input-orientation. It is based on available information about expected results and aims to improve accountability, efficiency and transparency. In this regard, providing and using evidence has an important value in the whole budget/programme cycle: programme assessment, budget development, implementation oversight, outcome monitoring and targeted evaluation.
Criminal justice	Jurisdiction	Strand 1: practice	Evidence is the means, sanctioned by rules, of ascertaining in a judicial proceeding the truth respecting a matter of fact. In this regard, fingerprints, blood, hair, skin, witness testimony and other items are understood as important incriminating types of evidence.
	Evidence-based criminology and criminal justice (crime prevention)	Strand 2: policy-making	Evidence is used in crime prevention and control, the treatment of prisoners and rehabilitating offenders based on quantitative and qualitative data and information.
Environment	Evidence-based environmental management and policy	Strand 2: policy-making	Advocates a more rational, rigorous and systematic approach to environmental management for supporting decision-making with the most reliable research findings on different environmental issues.
Source: Author			

Publications of the German Development Institute/ Deutsches Institut für Entwicklungspolitik (DIE)

Studies

98 Duguma, Mesay K., Michael Brüntrup, & Daniel Tsegai. (2017). *Policy options for improving drought resilience and its implication for food security: The cases of Ethiopia and Kenya* (87 pp.). ISBN 978-3-96021-048-1.

97 Reeg, Caroline. (2017). *Spatial development initiatives – potentials, challenges and policy lesson: With a specific outlook for inclusive agrocorridors in Sub-Sahara Africa* (176 pp.). ISBN 978-3-96021-048-1.

96 Hein, Jonas, & Britta Horstmann. (2017). *Aligning climate change mitigation and sustainable development under the UNFCCC: A critical assessment of the Clean Development Mechanism, the Green Climate Fund and REDD+* (131 pp.). ISBN 978-3-96021-043-6.

[Price: EUR 10.00; publications may be ordered from the DIE or through bookshops.]

Discussion Papers

7/2018 Baumann, Max-Otto. *Mission impossible? Country-level coordination in the UN development system* (41 pp.). ISBN 978-3-96021-063-4. DOI:10.23661/dp7.2018.

6/2018 Bergmann, Julian. *A bridge over troubled water? The Instrument contributing to Stability and Peace (IcSP) and the security-development nexus in EU external policy* (38 pp.). ISBN 978-3-96021-062-7. DOI:10.23661/dp6.2018.

5/2018 Furness, Mark. *Strategic policymaking and the German aid programme in the MENA region since the Arab uprisings* (33 pp.). ISBN 978-3-96021-061-0. DOI: 10.23661/dp5.2018.

4/2018 Minasyan, Anna. *Evidence-based allocation in global health: Lessons learned for Germany* (25 pp.). ISBN: 978-3-96021-060-3. DOI: 10.23661/dp4.2018.

3/2018 Thiele, Rainer, Maximilian Köster, Ikechukwu Okoli, & Friederike Rühmann. *African economic development: What role can the G20 Compact play?* (37 pp.). ISBN: 978-3-96021-059-7. DOI: 10.23661/dp3.2018.

2/2018 Ohnesorge, Jan. *A primer on blockchain technology and its potential for financial inclusion* (36 pp.). ISBN 978-3-96021-057-3. DOI:10.23661/dp2.2018.

1/2018 Ali, Murad. *Monitoring and evaluation in South-South Cooperation: The case of CPEC in Pakistan* (35 pp.). ISBN: 978-3-96021-058-0. DOI: 10.23661/dp1.2018.

30/2017 Martin-Shields, Charles. *State fragility as a cause of forced displacement: Identifying theoretical channels for empirical research* (21 pp.). ISBN 978-3-96021-055-9.

29/2017 Lundsgaarde, Erik. *The European Fund for Sustainable Development: Changing the game?* (33 pp.). ISBN 978-3-96021-054-2.

28/2017 Castillejo, Clare. *The EU Migration Partnership Framework: Time for a rethink?* (40 pp.). ISBN 978-3-96021-053-5.

27/2017 Hahn, Tina, & Georgeta Vidican-Auktor. *The effectiveness of Morocco's industrial policy in promoting a national automotive industry* (45 pp.). ISBN 978-3-96021-052-8.

[Price: EUR 6.00; publications may be ordered from the DIE or through bookshops.]

For a complete list of DIE publications:
www.die-gdi.de