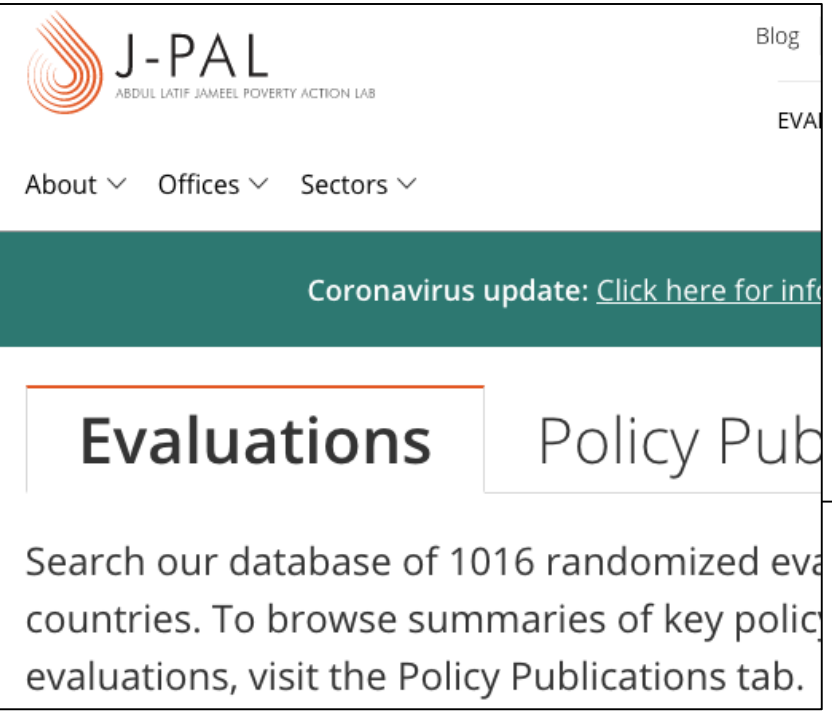
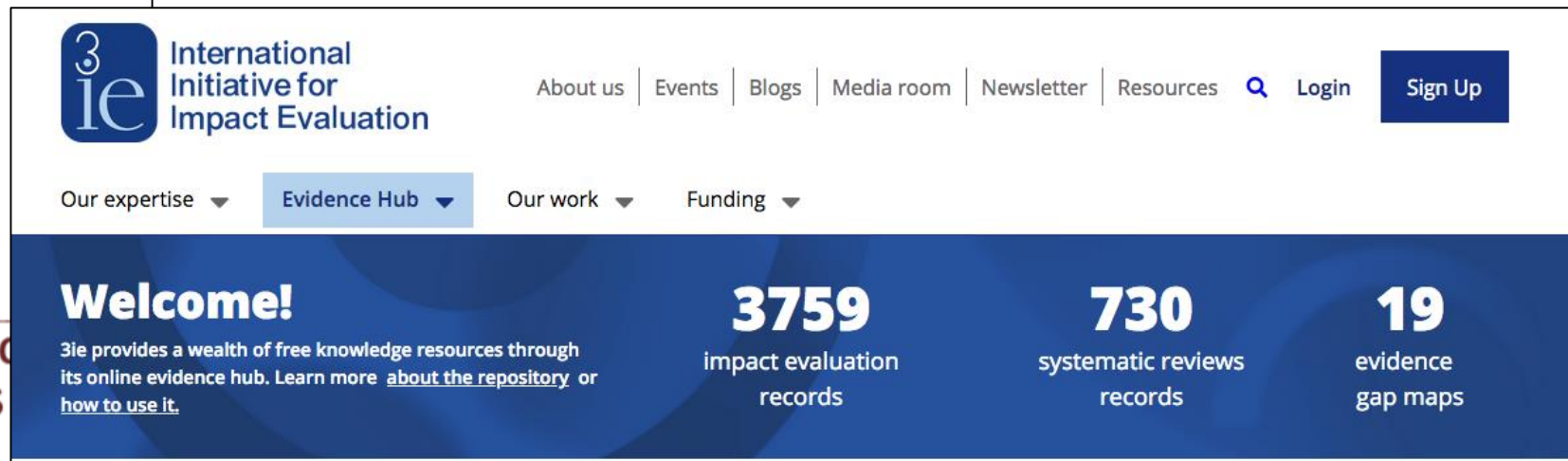
The Impact of Impact Evaluation

CEDIL Lecture, 17 June 2020

Richard Manning, Ian Goldman, Gonzalo Hernandez Licona





Outline

Core Question:

Nearly 15 years after the Centre for Global Development's report 'When will we ever learn?', are impact evaluation and related syntheses contributing to evidence generation and use in low- and middle-income countries?'

To that end, we address the following:

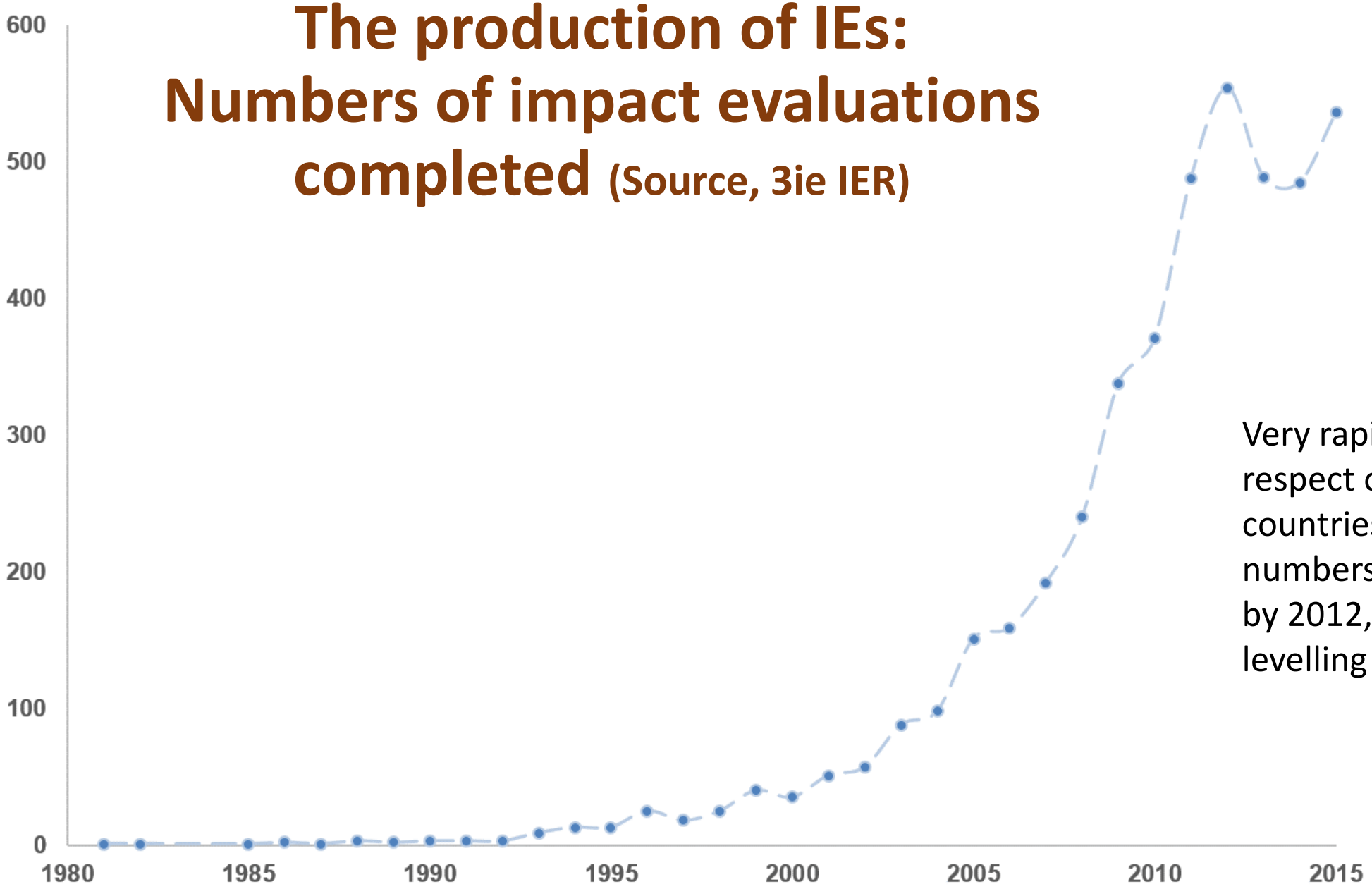
1. How has the generation of rigorous evidence developed in LMICs?
2. How to define and assess 'use'? What evidence of use?
3. What factors encourage use?
4. How transformative has the investment in impact evaluation and related syntheses in LMICs been, and where do we go from here?

Evidence

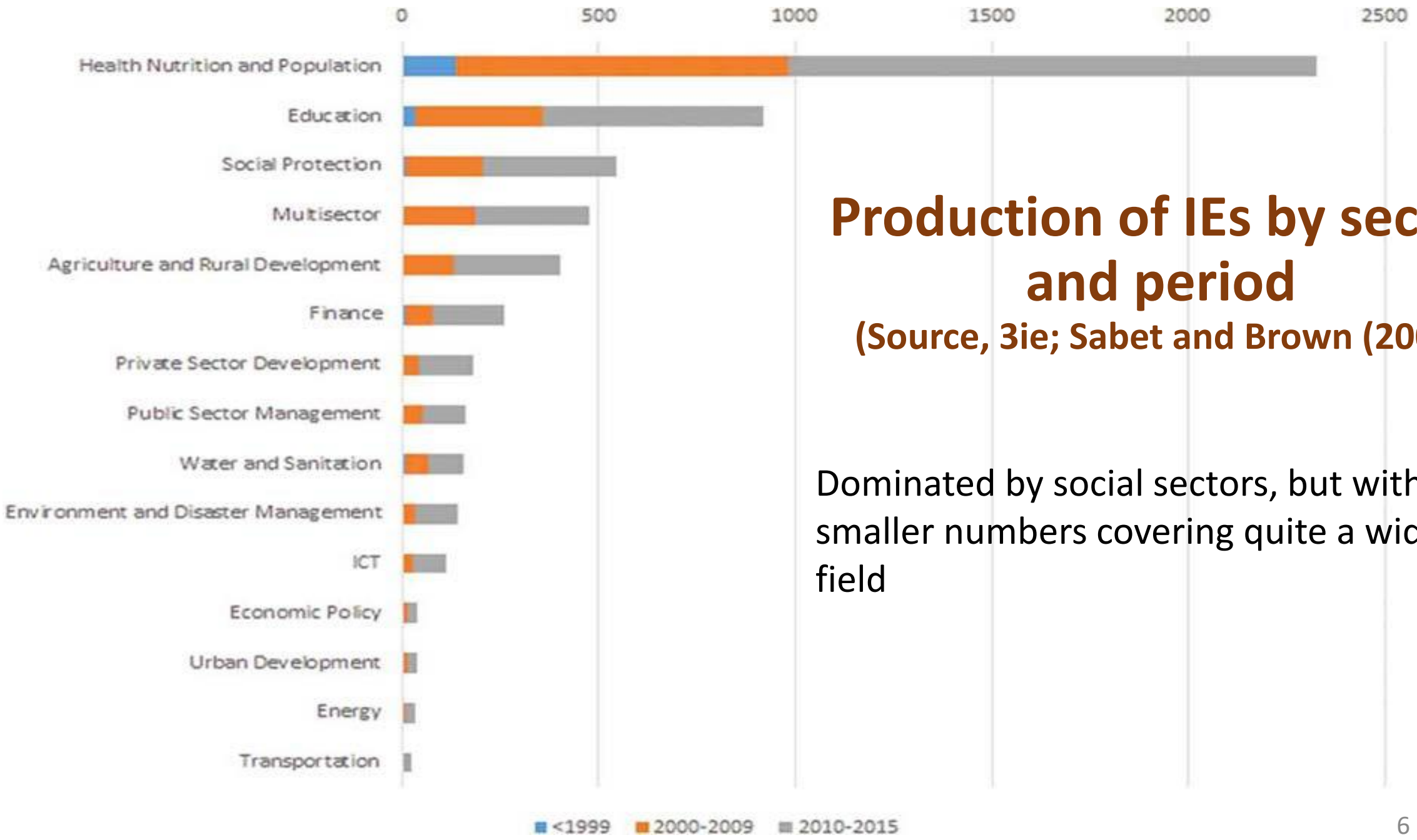
- Very few studies systematically assess **effects** of impact evaluations (IEs)
- Our approach was to gather relevant data through:
  - A) Drawing on existing databases (eg 3ie, J-PAL)
  - B) Structured questions to major funders, commissioners and suppliers of IEs and related syntheses
  - C) Five country studies (Mexico, Colombia, South Africa, Uganda, Philippines)
- Hope that our findings will encourage further discussion of how to make the best use for the benefit of LMICs of the tools developed for rigorous and relevant impact evaluation

The Production of Impact Evaluations

The production of IEs: Numbers of impact evaluations completed (Source, 3ie IER)



Very rapid increase in IEs in respect of developing countries from very low numbers to over 500 a year by 2012, but then a levelling off.



Production of IEs by sector and period

(Source, 3ie; Sabet and Brown (2008)

Dominated by social sectors, but with smaller numbers covering quite a wide field

The Production of Impact Evaluations

- Similar, but later, increase in SRs in international development - very few before 2008, over 100 published in 2016. (White 2019, quoting the 3ie database)
- RCTs become a staple of academic research, recognised with award of Nobel Prize to Bannerjee, Duflo and Kremer

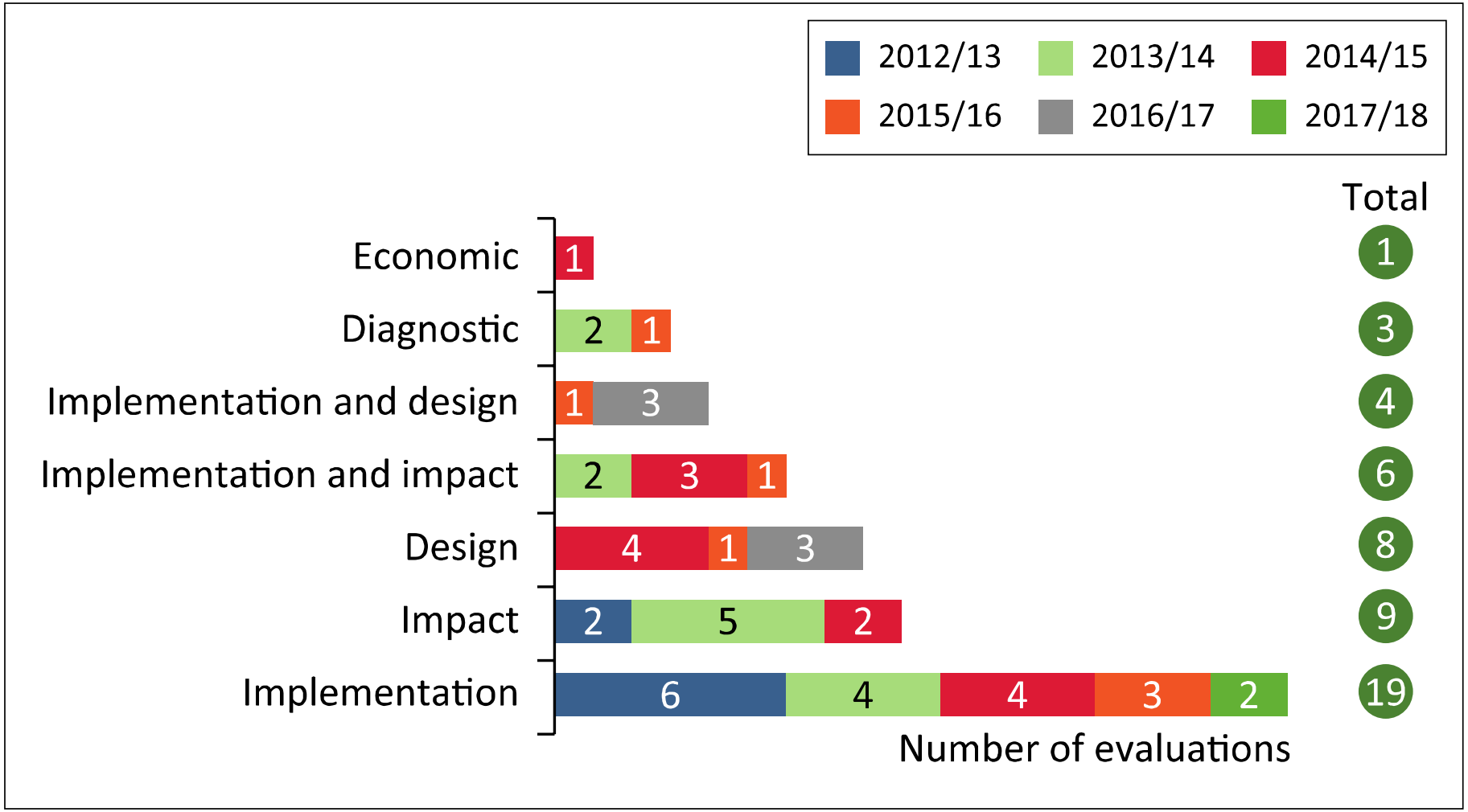
An evolving product line:
More rapid and policy-relevant IEs and synthesis products

Impact Evaluations

- *RCTs* still the main approach, but increasing use of robust *experimental and quasi-experimental approaches*, where randomization is not feasible
- An increasing focus on *evaluability*, and on identifying utility of IEs at the project design phase
- Coming back to *formative and process evaluations*, facilitating adaptive programming
- Development of less expensive and less time-consuming ways of carrying out IEs, while maintaining rigour.

Synthesis Products

- *SRs* being speeded up and more user-friendly
- *Other evidence reviews*, often commissioned by individual agencies, using a variety of protocols
- *Multi-country IEs* of issues, with findings brought together in 'policy insights' etc
- *Meta-analyses* of relevant groups of impact evaluations
- *Evidence [Gap] Maps* to identify areas for further policy attention

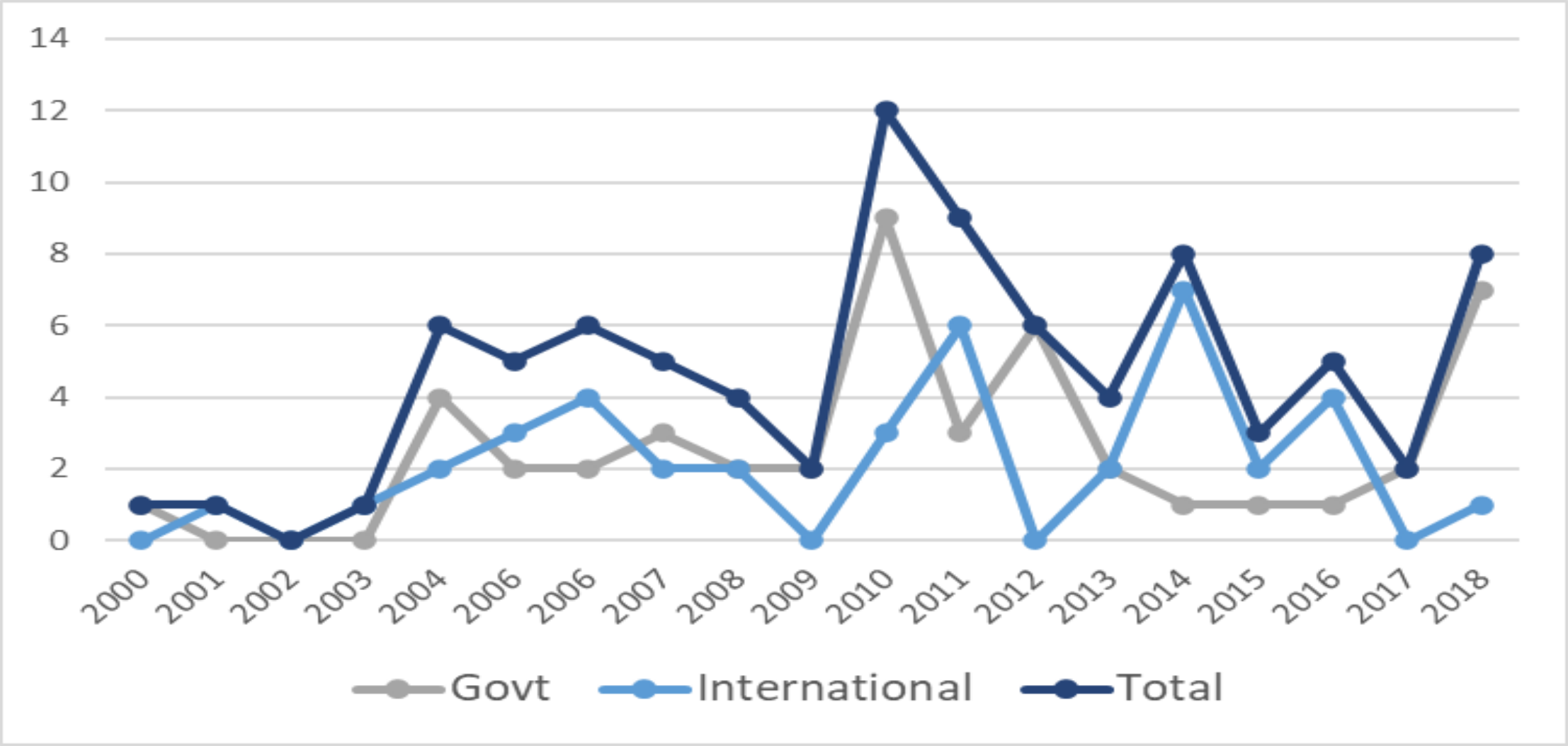


Evaluations by Type: South Africa, National Evaluation System

(note impact is not necessarily counterfactual based)

Source: Department of Planning, Monitoring and Evaluation (DPME), 2018c, *Report on the evaluation of the national evaluation system – Full report*, p. 82, Department of Planning, Monitoring and Evaluation, Pretoria

Number and Funding of IEs in Mexico, 2000-2018



Source: authors' arrangement based on information from CONEVAL, Crespo and Azuara Herrera (2017), and 3ie IER.

Notable features of the Production of Impact Evaluations in International Development

- IEs and Synthesis products still very small in relation to total evaluations:
  - Most donors fund very few IEs of their own programmes [MCC and USAID two exceptions];
  - In Mexico, 88 IEs carried out 2000-2018, but from 2007-2019 CONEVAL alone has co-ordinated over 2800 other kinds of evaluation
  - In Colombia and S Africa, there is a higher proportion of IEs in the total programme, but still modest
  - IEs in LMICs mainly internationally **funded**, even in an advanced country such as Mexico and predominantly so in a poorer country such as Uganda
- International funding from a dangerously narrow base, DFID in particular having been by far the largest funder of programmes through World Bank (DIME and SIEF) and major research programmes, and a major supporter of J-PAL, CEGA, 3ie etc; and the Gates Foundation also predominant among Foundations
- Donor-funded IEs are predominantly **commissioned** by donors or specialist intermediaries…….
- ……and very often still **led** by a relatively small number of 'Northern' institutions despite growing local capacity

Outline

1. How has the generation of rigorous evidence developed in LMICs?
2. How to define and assess ‘use’? What evidence of use?
3. What factors encourage use?
4. How transformative has the investment in impact evaluation and related syntheses in LMICs been, and where do we go from here?

Defining and Assessing Use: Applying Categories in the Literature (Johnson et al, 2009; Patton, 1997)

- Types of Use:
  - ***Instrumental use***: where someone has used evaluation knowledge directly
  - ***Conceptual use***: where people's understanding has been affected
  - ***Process use***: individual changes in thinking and behaviour and program or organizational changes in procedures and culture that occur…..as a result of the learning that occurs during the evaluation process
  - ***Symbolic use***: where a person uses the mere existence of the evaluation, rather than any aspect of its results (negatively to justify previous views, positively to elevate the status of the topic)

Applying these concepts appears most helpful in understanding the psychological process happening in evidence use.

For an understanding of what **type of use** happens, the more detailed categories in the next slide may be more helpful.

Types of Use:
A classification based on our findings

Largely instrumental

- Adjustments to programme activities
- Decisions to continue, expand, scale down, or cancel programmes
- Use to inform the design of new programmes
- Use to inform planning and budgeting processes

Largely conceptual/symbolic

- Influencing other related programmes, either externally or within the same implementing agency (probably conceptual);
- Use for policy dialogue and debate (conceptual, symbolic);
- Use to maintain political will (symbolic);
- Use to support and validate existing policies (possibly negative symbolic);
- Use of large bodies of evidence (so, particularly SRs) to inform wider thinking (conceptual);

Largely process

- Use to justify expanding the use of M&E and IEs (instrumental, conceptual)
- Improving the culture of evidence use (conceptual)

Evidence of Use

- Almost no counterfactual based evidence and just a couple of independent evaluations of IE use by agencies
- Based on what we could find from all sources, we conclude:
  - Under ‘Largely Instrumental’: quite a lot of programme-specific use, including some examples of improving follow-on design; not much on planning and budgeting processes
  - Under ‘Largely conceptual/symbolic’: some high-profile examples (eg cash transfers, micro credit); but less than one would have hoped
  - Under ‘Largely process’: broad evidence of higher interest in IEs and M&E generally in implementing agencies exposed to IEs; anecdotal evidence of IEs ratcheting up quality and rigour of other types of evaluative work
- More details in slide pack

What do we know about Use? (1)

Evidence from Experimental approaches

- Little counterfactual-based evidence. Brazilian Municipalities study (Hjort et al, 2019) does however represent one recent experimental approach. It found that:
  - Mayors and other municipal officials were willing to pay to learn the results of IEs (particularly with large samples), and would update their beliefs when informed of the findings
  - Informing mayors about research on a simple and effective policy (reminder letters for taxpayers) increased the probability that their municipality implemented the policy by 10 percentage points

What do we know about Use? (2)

Independent Evaluations: World Bank IEG (Ramirez et al, 2012) and IDB Office of Evaluation and Oversight (Crespo and Herrera, 2017).

Instrumental Use:

- WB: 'modest'. <half of completed World Bank IEs were mentioned in project completion documents to demonstrate project impact.
- IDB: 56% of interviewees stated that IEs had had influenced or were expected to influence policymaking. For ongoing IEs was this rose to 83%.

Conceptual Use:

- Both institutions noted value of IEs for policy dialogue
- Institutional strategies had benefited where there was a large body of evidence, such as education and social protection , but not always systematically
- IDB noted that IE evidence had discouraged investment in 'One Laptop per Child' program.

Process Use (WB):

- A third of completed IEs considered to help improve capacity in conducting or analysis of IEs.
- In 65% of follow-on projects an IE was planned for interventions at the appraisal stage. Five IEs had contributed to encouraging a more evidence-based policy-making culture.

Symbolic Use:

- Both observed value of PROGRESA IE for supporting other cash transfer programs.

What do we know about Use? (3a)

Instrumental Use 1: Surveys and self-reporting by Agencies

- 66% of survey respondents agreed World Bank's IEs under 'DIME' informed programme/policy design, and 82% they helped rationalize existing designs
- 3ie observed 30 changes to policy or programme design in a sample of 86 IEs and related projects completed between 2013 and 2018
- J-PAL found that providing identification cards to beneficiary households improved access to Indonesia's national rice subsidy programme; and this finding contributed to the GoI's decision to scale up such cards for a range of programmes targeted at the poorest households.

What do we know about Use? (3a)

Instrumental Use 2: Evidence from Country Studies

Mexico: Milk Programme IE: Government widened distribution of fortified milk to other programmes, such as the Rural Supply Programme; Food Programme IE: cash support made conditional on attending nutrition talks and receiving nutritional supplements and overall health support; Cement Floor Programme IE: budget increased. The scale of use from CONEVAL evaluations (NOT just IEs) is evident from [next slide]

Colombia: Familias en Acción. Programme re-structured after IEs showed increased school attendance but not better educational outcomes, and labour market impact only in rural areas

S Africa: Child support grant IE key in supporting extension of age at which children eligible to 18 years, and to counteract the belief that the grant encouraged teenage pregnancy; Youth Employment tax incentive IE contributed to decision to expand the scheme. But pre-school IE recommending against additional year until quality issues were addressed not adopted by gov.

Uganda: Youth Livelihood Programme evaluation resulted in a 30% increase in allocation of funds to beneficiaries. A SR of food fortification resulted in the policy being adopted nationally.

Philippines: Successive IEs have led to scaling up and institutionalization of CCT programme; IEs have improved implementation of Community Driven Development Programme, and triggered a greater focus on skills training in a Students' Employment programme.

What do we know about Use? (3b)

Conceptual Use 1: Surveys and self-reporting by Agencies

- Over two-thirds of government respondents to DIME's survey said that they used the evidence or data from the IE to guide the design of other projects
- The 3ie study shows 27 cases where IEs and associated products informed the design of other programmes, including internationally
- Synthesis products (not just SRs) have led to wider attention to, eg, cash transfers, early years education, HIV self-testing, free distribution of bed-nets; and have raised questions about micro-credit and cookstoves.
- However, outside medicine, few international policy communities themselves seem to have taken a lead in encouraging or commissioning, let alone funding synthesis work on issues of significance to them.

What do we know about Use? (3b)

Conceptual Use 2: Evidence from Country Studies

Mexico: CONEVAL is also responsible for measuring poverty at national and state level. This generates government demand for evidence of the effectiveness of programmes in reducing multidimensional poverty.

Colombia: IE methodologies have spread to a wider range of public policy sectors such as retirement funds, health, education, transport, culture, agriculture, and housing.

S Africa: Respondents from the Dept of Human Settlements noted that they found evaluations helpful as a reflective experience, but not currently as a decision-making exercise.

Uganda: Family Planning IE enhanced ongoing debates in the country about the content of the sex education provided to young people in schools. Universal Primary Education IE informed the design of the new Education and Sports Strategic Plan and the formulation of the Theory of Change.

What do we know about Use? (3c)

Process Use (1): Surveys and self-reporting by Agencies

- 3ie study found 28 examples of IEs improving the culture of evaluation evidence (eg leading to further evaluative work, or building skills and interest in the implementing agency)
- Several respondents from donor agencies reported that IEs had improved the rigour and quality of other evaluative work in their agency and one referred to a ‘far more nuanced discourse and understanding about impact and what types of evidence can be generated by what types of evaluation methods’

What do we know about Use? (3c)

Process Use 2: Evidence from Country Studies

Mexico: In 2001, Congress required external evaluations for every social programme, arising from IE of Progresa. This led to creation of CONEVAL in 2006 as independent institution to evaluate social policy/ programmes and, with Ministry of Finance, to build one of the most thorough-going M&E systems

Colombia: Sinergia founded in 1994, gained credibility in part from donor-funded IEs, and from 2008 more resources were allocated to enable it to conduct evaluations itself.

S Africa: the evaluation of the Evaluation System found 'unintended benefits' reported by departments and provinces included: (1) improved strategic vision as a result of using theories of change; (2) use of 'good practice' in internal research after exposure to external evaluations; (3) enhanced use of evaluative thinking; and (4) the need to harmonize learning across structures.

Uganda: Government focus on M&E included a Government Evaluation Facility (mostly financing process evaluations) and 3ie funding for 3 IEs. Uganda also a pioneer in Africa in use of SRs through The Africa Centre for Systematic Reviews and Knowledge Translation at Makerere University, and the African Centre for Rapid Evidence Synthesis (ACRES).

Philippines: In 2015, the Australian Government provided AUS$2.8m through 3ie for conduct of rigorous IEs and capacity-building workshops. A steering group, chaired by the DG of the National Economic and Development Authority, mobilizes the rest of government to identify IE topics and then decides how to allocate the funds. This has built ownership and capacity in government.

What do we know about Use? (3d)

Symbolic Use

This category is less well documented.

- Conceivable that the critical mass of (positive) IEs of cash transfers has significantly helped spread this approach to social programmes. Indeed, first IE of Progresa could be seen as a 'signal' of the worth of that programme, and led to interest in similar high-profile studies in other sectors
- A national evaluation of nutrition interventions for children under five in South Africa raised the profile of the issue of child malnutrition
- The negative aspect of symbolic use is most common where IEs are used to justify an existing policy position, sometimes referred to as 'policy-based evidence'. The IDB evaluation included a client survey, where 52% of interviewees saw IE primarily as a tool to adjust policy, while 20% identified it as an instrument to support and validate existing policies

Outline

1. How has the generation of rigorous evidence developed in LMICs?
2. How to define and assess ‘use’? What evidence of use?
3. What factors encourage use?
4. How transformative has the investment in impact evaluation and related syntheses in LMICs been, and where do we go from here?

Factors Affecting Use: What Works?

(Systematic Review of Research Uptake, Langer et al, 2016)

Effective

- Facilitating ***access*** to research evidence, **IF** the intervention simultaneously tries to enhance decision-makers' *opportunities* and *motivation* to use evidence (reliable evidence).
- Building ***decision-makers' skills*** to access and make sense of evidence **IF** the intervention simultaneously tries to enhance both *capability* and *motivation* to use research evidence (reliable evidence).
- Fostering ***changes to decision-making structures and processes*** by formalising and embedding one or more of the other mechanisms of change within existing structures and processes (such as evidence-on-demand services integrating push, user-pull and exchange approaches) (cautious evidence)

Ineffective:

- A ***passive approach to communicating evidence***, (such as simple dissemination tools) (reliable evidence).
- *A* ***passive approach to building evidence-informed decision-making (EIDM) skills*** without active educational components) (cautious evidence).
- Skill-building interventions applied at a ***low intensity*** (cautious evidence).
- Overall, ***unstructured interaction and collaboration*** between decision-makers and researchers tended to have a lower likelihood of success. However, clearly defined, light-touch approaches to facilitating interaction between researchers and decision-makers were effective in increasing intermediate outcomes

Features specific to IEs Funded by International Actors

- **Implications of *Donor Funding***
  - If funded for purpose of increasing knowledge of some significant issue/intervention, can be very positive, though may be questions of buy-in from other key stakeholders, and evidence of effective use for internal learning of agencies very mixed
  - If funded for domestic accountability, wider use of findings less likely
  - Links to local 'evidence systems' not often seen as a priority
- **Implications of *Commissioning by Donors or Agencies***
  - Some anecdotal evidence that commissioning by LMICs would help embed local ownership
  - Similar argument for more commissioning by international policy communities
  - Scope for more engagement of clients even while taking advantage of experienced commissioning systems
- **Implications of *'Northern'- led Production***
  - More locally-led IEs might have more local credibility, as well as more in-depth knowledge of local context
  - Some risk of less independence? Quality obviously vital, but growing local capacity
  - Evidence of some greater share of local leadership, but a long way to go
- **More focus on instrumental use**
  - But also examples eg of 3ie thematic windows where contributing to evidence in the sector (conceptual use)
  - Less opportunities for process use in country, but is within the donors

Factors Affecting Use: What we found

1. from the Agencies

1. There is **no simple linear relationship** between producing evidence and having it used. Understanding the context of the intervention is essential.
2. There may be **different paths for evidence** that is of direct relevance to a programme being designed or scaled up (e.g. an IE of a pilot phase) and evaluations of established programmes,. The former is usually easier to implement. For the latter, the accretion of knowledge (eg through synthesis products) may be more effective in encouraging reconsideration of the intervention.
3. **Timeliness** and relevance to the context are very important in determining the likelihood of use
4. Most institutions that finance development-oriented IEs have given increasing weight to building effective **links between researchers and policymakers** (see next slide).
5. While '**champions**' within governments can be very influential in translating the advice of researchers into actionable policy, most funders recognize that it is usually inadequate to rely too much on this element.

Factors Affecting Use: What we found

Good practice in engaging with policy-makers

Examples:

- Responding to policy-makers ready to take central role (eg Tamil Nadu and J-PAL)
- Early discussion between policy-makers and researchers (eg DIME workshops)
- Engaging with policy-makers throughout process (3ie and others)
- Involving other stakeholders (eg SIEF with journalists)
- Supporting implementation [J-PAL's Innovation in Government Initiative funds technical assistance to adapt, pilot, and scale evidence-informed innovations that have been previously undergone randomized evaluation and found effective]
- Capacity building linked to IE (Eg CEGA/East Africa)

- As often, good practice not yet general practice – and of course interventions need to be context-specific

Factors Affecting Use: What we found

2. from the country studies: Facilitators

1. **Political will** in government, at both political/senior official and project levels
2. Existence of a government-backed **M&E or evaluation strategy**, with dedicated funding (and also a basis for using international funding), and resources to stimulate supply of evaluations and work with sector depts on implementation
3. **Link to central government** functions eg:
    1. Mexico, the Budgetary Consideration Report, for Congress, contains assessments of all social development programmes, with budget recommendations;
    2. Colombia, positioning of Sinergia in Dept for National Planning;
    3. SA, integration of IEs in the government-wide M&E system under Dept for Planning, M&E;
    4. Uganda, commitment by Treasury to use results from evaluations to inform resource allocation;
    5. Philippines, the central role of the National Economic and Development Authority)
4. High degree of **transparency**, and the involvement of the legislature

Factors Affecting Use: What we found

2. from the country studies (Facilitators, continued)

5. Ensuring **buy-in from the eventual owners of the study** right from the evaluation design, and working with (and encouraging) champions in the departments concerned
6. Being ready to **explain/discuss the differences** between impact evaluation, performance monitoring, and other types of evaluation and their intended purposes and outcomes, using eg formative evaluation where it is appropriate
7. Developing **recommendations in consultation with stakeholders** and maximizing knowledge-sharing through wider stakeholder validation and presentation of study outputs
8. Ensuring the **quality** of IEs
9. **Local centres of excellence** in the production of IEs
10. A clear **system for addressing recommendations**, with regular progress reports on them (eg in Mexico and SA the use of improvement plans following evaluations. The follow-up of recommendations makes it possible to systematically monitor use.)

Factors Affecting Use: What we found

2. from the country studies: Barriers

Barriers to application of IEs

- **Limited awareness of IEs** outside the health sector
- The **time and cost** needed to undertake evaluations, especially IEs: need for additional funding
- Lack of available and accessible quality **baseline data**
- IEs **not commissioned early enough**
- **Complexities** in programmes which make IEs difficult
- Limited **supply of local evaluators** able to lead counterfactual IEs

Barriers to use of IE findings

- Failure to understand information **stakeholders need** when designing studies/not enough attention paid to how the evaluation will be used by government, as opposed to evaluation methodology
- A **culture**, at least in some countries, that does not accept the kind of criticisms that evaluations inevitably present
- **Inadequate institutionalization of findings**/ systems for follow up of recommendations

Outline

1. How has the generation of rigorous evidence developed in LMICs?
2. How to define and assess ‘use’? What evidence of use?
3. What factors encourage use?
4. How transformative has the investment in impact evaluation and related syntheses in LMICs been, and where do we go from here?

Have we met the Challenge of the CGD Report, 'When will we ever Learn?' (1)

The report argued for a new collective push to promote rigorous evaluations of impact. It concluded:

"Will we really know more in 10 years? […] The international community could be in one of two situations.

- We could be as we are today, bemoaning the lack of knowledge about what really works and groping for new ideas and approaches to tackle the critical challenges of strengthening health systems, improving learning outcomes, and combating the scourge of extreme poverty.
- Or we could be far better able to productively use the resources for development, based on an expanded base of evidence about the effectiveness of social development strategies.

- Which of those situations comes to pass has much to do with the decisions that leaders in developing country governments, NGOs, and development agencies make over the next couple of years about conducting impact evaluations"
- Has this challenge been met?

Have we met the Challenge of the CGD Report, 'When will we ever Learn?' (2)

Conclusion

1. Progress in rigorous evaluation is undeniable: We are almost certainly *somewhat* 'better able to productively use the resources for development, based on an expanded base of evidence about the effectiveness of social development strategies'
2. But it seems doubtful that the expansion of IEs and associated products has been as transformational as hoped by the authors of the 'When Will We Ever Learn?' report – notably in getting beyond the project level
3. The challenge of complexity:
   - LMICs are faced with highly complex problems of sustainable growth, societal change, personal wellbeing, and the development of competent institutions. Many deep-seated issues are not readily amenable to standard IE techniques, despite much progress.
   - The pathway to policy influence is seldom straightforward. Interests often inhibit the changes that evidence recommends. LMICs are not at all immune to rhetoric about 'post-truth' and rejection of 'experts'.
4. Not a time to give up on rigorous evidence!

Where do we go from here? (1)

- Increased funding over last 10–15 years made it possible to **experiment** with many different ways of bringing researchers and policymakers together in constructive ways, both at country level and in international policy communities, engaging not just governments but also legislatures, civil society, and all forms of media.
- Also now **better understanding** of how IEs fit within a wider range of tools available to support policymakers with timely and grounded evidence.
- A good time to reflect on how to **promote better evidence systems** and **better use** of the evidence that IEs and associated syntheses provide. Much scope for learning between countries, and indeed donor agencies.

Where do we go from here? (2)

- Greater **local ownership** of IEs is highly desirable.
- Donor finance and commissioning has been key, but our sense is that — as with other forms of evaluation — **a more balanced pattern of finance, commissioning and supply of IEs is needed** if IEs are to become a more accepted part of national evidence systems.
- Important that scarce **donor funding** is responsive to priority concerns of countries and of the international community.
- Appears to be reduced appetite by donors to fund **underlying public goods** such as repositories of IEs. *As much knowledge is gained by accretion, this would be unfortunate.*
- We hope that our survey will encourage **greater reflection** by all parties on the lessons to be drawn on good practice from the growing body of evidence tools, and from IEs and related synthesis products in particular.

Thank you

Richard Manning

richardgmanning@btinternet.com

Ian Goldman

Ian.goldman@wits.ac.za

@iangoldmansa

Gonzalo Hernandez Licona

ghlicona@gmail.com

@GHLicona

Link for full paper:
https://www.wider.unu.edu/sites/default/files/Publications/Working-paper/PDF/wp2020-20.pdf