



# IEO Background Paper



## Strategy Design in Evaluating IMF Surveillance Activity

*Paul Duignan and Nils Bjorksten*

## **IEO Background Paper**

Independent Evaluation Office

Strategy Design in Evaluating IMF Surveillance Activity

Prepared by Paul Duignan and Nils Bjorksten

June 6, 2005

### **Abstract**

For both operational and accountability reasons, the IMF continually evaluates its surveillance activities. Doing this involves making nontrivial evaluation strategy choices, such as finding the right balance between internal and external evaluation components, deciding at which nodes of the results chain that evaluation efforts should be focused, and deciding how to build on the information that is collected. We set out a methodology (REMLogic) for formulating an overall strategy for evaluating surveillance that aims to deliver the most useful information to decision makers at the highest feasible level of causal attribution to help improve the overall effectiveness of surveillance. We also present some concrete examples that help to make the proposed approach operational.

<p>The views expressed in this Background Paper are those of the author(s) and do not necessarily represent those of the IMF, IMF policy or the IEO. Background Papers report analyses related to the work of the IEO and are published to elicit comments and to further debate.</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Keywords: IMF Evaluation, REMLogic, Strategy Design, Surveillance

Authors' E-Mail Addresses: [paul@parkerduignan.com](mailto:paul@parkerduignan.com) and [nbjorksten@imf.org](mailto:nbjorksten@imf.org).

## I. INTRODUCTION

Evaluation of IMF surveillance activity is as challenging as it is important. Since the overall costs of conducting IMF surveillance are high, at more than one-quarter of the IMF's administrative budget of over \$900 million, there is much at stake in determining how much IMF surveillance is enough, and in trying to make this surveillance activity as efficient and effective as possible.<sup>1</sup>

We set out an approach to developing an evaluation strategy that takes into account the complementary roles of internal and external evaluations of surveillance, and in particular we consider the specific evaluation aims that are reasonably assigned to the internal Policy Development and Review (PDR) department and to the IMF's Independent Evaluation Office (IEO), respectively. We identify the danger of wasting resources on misleading "pseudo-outcome evaluations" which purport to measure attributably surveillance's effectiveness in changing high-level outcomes, in cases where such definitive overall evaluation, no matter how desirable, is simply not methodologically feasible.

Finally, we present the general principles and describe a specific technique (the REMLogic methodology), which aims to ensure that evaluation resources are most usefully allocated for the purpose of obtaining the most helpful strategic evaluation information for decision makers. This methodology attempts to push causal attribution as high as possible towards final outcomes in order to make sure that the institution is not avoiding the key attributional questions stakeholders want answered, but avoids obfuscating pseudo-outcome evaluations by not pushing beyond what is technically feasible and affordable. By including indicator sets as an essential building block in its approach to evaluation planning, this approach is able to integrate well with existing performance measurement and monitoring initiatives.

The remainder of the paper is structured as follows. Section II defines what we mean by surveillance and presents an overview of the role and logic of evaluation at the IMF, including the respective functions of internal and external reviewers. We present two alternative paradigms for evaluating surveillance and identify the risks of trying to push assessments of overall effectiveness beyond what is feasible. Section III describes a 12 step methodology for determining an evaluation strategy for IMF surveillance, which we refer to as the REMLogic approach. Section IV illustrates how the entire approach can be made operational in the context of evaluating multilateral surveillance, elaborates on available methodological choices as to collecting evaluation evidence, and describes some design

---

<sup>1</sup> Budgeted expenditure on surveillance was just over \$240 million in FY2006, more than a quarter of the total gross administrative budget of the IMF. This includes the cost of well over 500 staff years, as well as overhead and travel. It does not include the costs of country officials' time and resources in each of the countries being examined.

issues to consider in combining the different methodologies most effectively. Section V concludes.

## II. EVALUATING SURVEILLANCE

Surveillance by the IMF refers generically to all activities engaged in for the purpose of enabling the IMF to (i) oversee the international monetary system to ensure its effective operation, and (ii) to oversee members' compliance with the obligations specified under Article IV of the IMF's Articles of Agreement. These activities include all analysis of, and advice concerning, member countries' current economic situations and prospects, with the ultimate objective of furthering the well-being of all IMF members, individually and collectively, as defined in Article I of the Articles of Agreement.

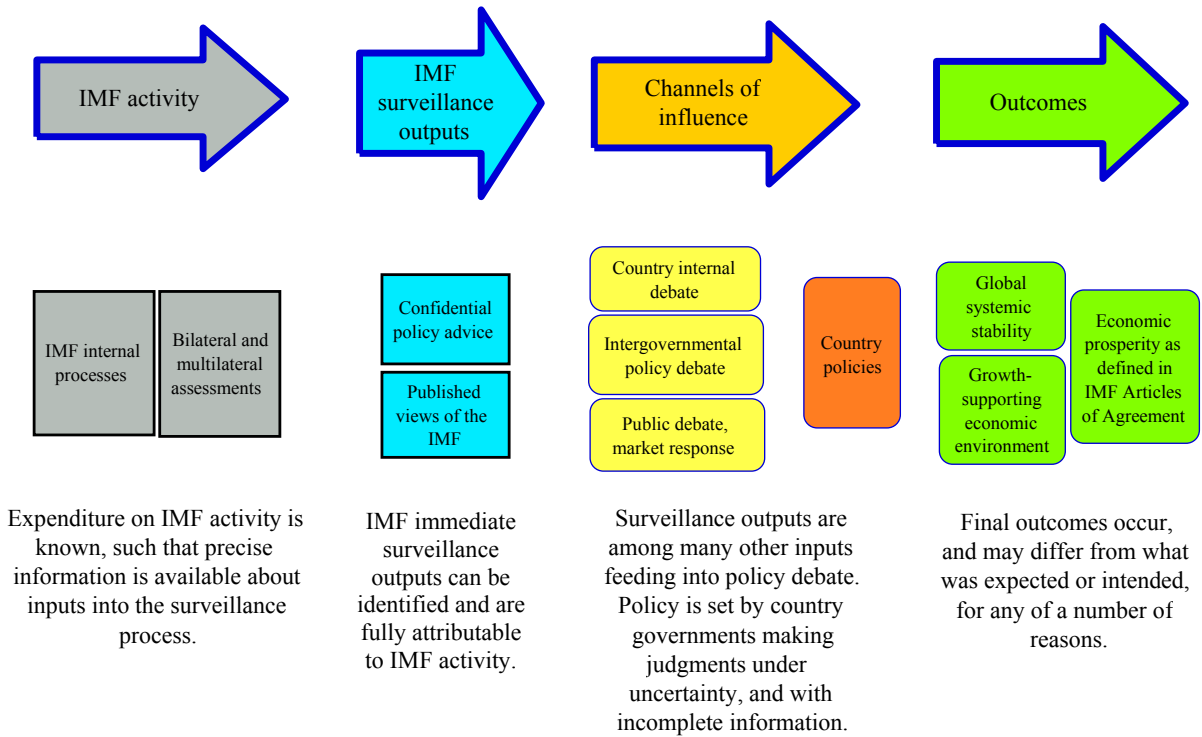
While the objectives of surveillance have remained unchanged, at least since the Articles of Agreement in 1978, it is well-recognized that the modalities of surveillance, as well as the channels through which surveillance affects outcomes, have evolved with changes in the world economy.<sup>2</sup>

Figure 1 below presents the main components of IMF surveillance today, together with their downstream links to outcomes. The figure shows the rough progression from inputs to IMF surveillance outputs, which in turn enter the institutional channels of influence and eventually feed into the final outcomes that surveillance activity is intended to have an impact on. Elements under each heading are tightly connected and interlinked, and, as reflected in the changes in modalities of surveillance mentioned above, many of the connections within and between groups of boxes under each heading continually evolve over time.

---

<sup>2</sup> The present modalities of IMF surveillance were concisely described by the IMF Executive Board in its Public Information Notice (PIN) 04/95 of August 24, 2004: "The IMF fulfils this [surveillance] mandate through bilateral, regional, and multilateral surveillance. In accordance with Article IV of its Articles of Agreement, the main instrument of bilateral surveillance is consultations, normally held every year, with each of the Fund's members. These consultations are complemented with regular analysis of economic and financial data provided by members and, as needed, informal contacts between the Fund and national authorities. At the regional level, the IMF holds regular discussions with economic institutions of currency unions and participates in the activities of regional bodies. The pillars of the Fund's multilateral surveillance are the World Economic Outlook report and the Global Financial Stability Report, which are produced twice a year. The reports are complemented by more frequent, informal reviews of global economic and market developments."

Figure 1. Components of IMF surveillance: inputs, outputs, channels of influence and intended outcomes



### A. The Role and Logic of Evaluation

In order to adapt the modalities of surveillance to a continually evolving institutional environment, a system of periodic evaluation of IMF activity is essential. Likewise, because of the costliness to its members of the IMF’s surveillance activity, evaluation is also required for public accountability purposes. It is, therefore, not surprising that surveillance is one of the most evaluated functions of the IMF, with the IMF’s Policy Development and Review (PDR) Department conducting a comprehensive surveillance review every other year.<sup>3</sup>

As we take stock of previous and ongoing evaluations of surveillance, it is useful to give thought to developing a general evaluation strategy for IMF surveillance that will both further illuminate the process and help in future high level strategic decision making.

A comprehensive program of evaluation of surveillance has roles for both internal and external evaluators. PDR’s Biennial Surveillance Review is in practice an internal self-evaluation exercise by the IMF, and as such it is well suited for the purpose of improving operational aspects of surveillance. But, partly because surveillance practice at the IMF

<sup>3</sup> The report prepared by PDR is the basis for the IMF Board’s Biennial Surveillance Review. Between 1977 and 1988, the review was on an annual basis.

includes a central coordinating role for the very department charged with conducting this review, and partly because the staff conducting the review reports to IMF management, it follows that PDR is not institutionally positioned to deliver a truly independent assessment of IMF surveillance activity. Self-evaluation is simply not well suited for the purposes of providing accountability and full transparency into what is, in fact, a costly and complex set of interconnected activities. In order to counter the vulnerabilities of internal evaluation, an external evaluation exercise was commissioned in 1999,<sup>4</sup> and subsequently the Independent Evaluation Office (IEO) was established. A system of performance indicators is also currently being discussed in response to calls for the IMF to develop better approaches for monitoring the effectiveness of surveillance.

The IEO has recently begun an evaluation of some aspects of IMF surveillance. In separate evaluation exercises, the IEO is currently examining the Financial Sector Assessment Program (FSAP), as well as multilateral surveillance.<sup>5</sup> At some point in the future, the IEO may also undertake an evaluation of bilateral surveillance of systemically important countries. Because of its independence from the management of the IMF, the IEO is better positioned to provide the wider public with a transparent and dispassionate description of the surveillance process at the IMF, including the logic and practice as it has evolved in recent years. It is also positioned to assess, from an accountability perspective, the performance of the surveillance framework as currently implemented at the IMF, and to review those aspects of surveillance that call for IMF decisions at strategic rather than at operational levels.

## **B. Choice of an Evaluation Paradigm**

Designing evaluations involves making some important decisions. As a general principle, the choice of an evaluation paradigm should depend on both the evaluation needs that the evaluation is intended to satisfy, and on feasibility and cost constraints. In the case of IMF surveillance, both aspects bear careful consideration.

### **The Black Box paradigm**

At the simplest level, the IMF Executive Board (and member country governments), management, staff, other stakeholders and the general public are understandably interested in establishing the overall effectiveness of IMF surveillance activities. Suppose for a moment that this were the only question that evaluators were asked to address, i.e., that the sole purpose of the evaluation were to prove an attributable causal relationship between IMF surveillance activity and high-level outcomes, with no interest in delving into the particular channels through which the effects may operate. This methodological approach is known as a “black box” evaluation, because the specific linkages between IMF surveillance activity and

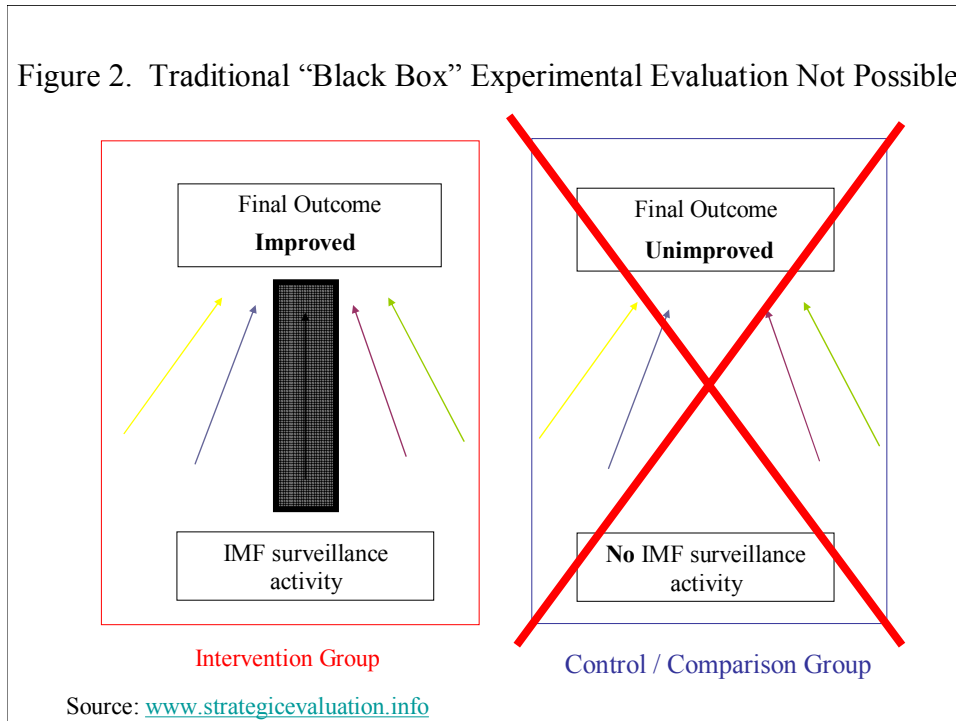
---

<sup>4</sup> External Evaluation of Surveillance Report: Report by a Group of Independent Experts, IMF, 1999.

<sup>5</sup> See [www.imf.org/ieo](http://www.imf.org/ieo) for issues papers/terms of reference for these evaluations.

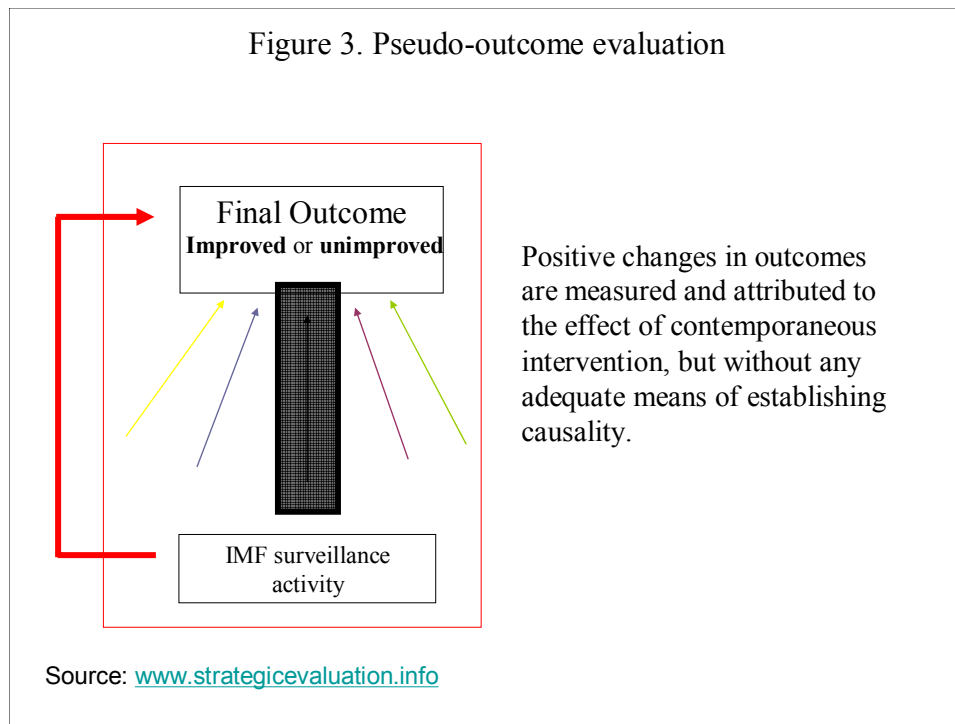
final outcomes are hidden in a “black box” and thus what happens to intermediate outcomes remains outside the scope of the evaluation.

If this approach were feasible, it would potentially provide very useful information for IMF and stakeholder decision making. However, in order to use the black box paradigm for establishing an attributable **causal** relationship, the traditional approach would be for surveillance to be undertaken with one group of countries, and the outcomes for these would be compared to those for an equivalent group not subject to IMF surveillance (i.e., a control group).<sup>6</sup> This is a very useful paradigm for conducting outcome evaluations (e.g., medical experiments), where, as with surveillance, the end outcomes depend on a variety of complex factors which need to be controlled for. **But the black box high level outcome evaluation approach cannot work for IMF surveillance because no control group of countries exists which can be completely detached from any IMF influence** (see Figure 2).



<sup>6</sup> There is currently debate in the evaluation discipline regarding the traditional black box experimental paradigm as the only or best way of establishing causation in all cases. There are some alternative methods proposed for establishing causation, which can be used in some cases and aspects of these can be drawn on in the design of IMF surveillance evaluation. However, given the nature of surveillance activity and the multiple other factors affecting final outcomes, it is unlikely that any of these alternatives will be able to establish a causal connection between IMF surveillance and final outcomes in the same way that such causal connections are established for “easier to evaluate” programs such as the effectiveness of pharmaceuticals in medicine.

In cases like this, it is a serious error to allow the black box paradigm to become the basis for evaluation planning. If this paradigm is still allowed to dominate evaluation planning where no control group exists, the result is that effort and resources can be wasted on *pseudo-outcome evaluations*. Such evaluations pretend, and perhaps even lead certain stakeholders into believing, that final outcomes can unambiguously be attributed to IMF surveillance activity in spite of the presence of many other important influences that have not been controlled for. In actual fact, what is usually done in such cases is simply measuring changes in final outcomes and drawing contestable conclusions about the contribution of surveillance to these changes. The process is illustrated in Figure 3.



Is the IMF at risk of wasting resources on pseudo-outcome evaluations of surveillance’s overall high level effectiveness? Looking at IMF Executive Board discussions of evaluation of effectiveness,<sup>7</sup> there is a desire to “make further progress in assessing effectiveness of surveillance,” and “setting monitorable strategic objectives” as part of ongoing monitoring. The broader discussion nevertheless leaves unclear as to how ambitiously this “assessing effectiveness of surveillance” is best viewed. If this means establishing attributable causation between IMF activity and the highest level of outcomes, this may not be possible and a futile attempt to push outcome evaluation beyond its feasible level will lead to pseudo-outcome

<sup>7</sup> For instance Public Information Notice (PIN) No. 03/50, April 10, 2003 and PIN No. 03/116, September 10, 2003.



evaluations.<sup>8</sup> But if it means pushing the attempt at attribution as far up as possible towards high level outcomes, then this is an entirely feasible approach.

Evaluation decisions then center around selecting the most practically useful and achievable evaluation projects for ongoing improvement of IMF strategy in regard to surveillance including, but not restricted to, attribution at the highest feasible level. The results of such evaluation projects would then be used to inform the ongoing development of IMF surveillance strategy going forward.

### **The “hard to evaluate” paradigm**

While the black box approach has been ruled out as a viable paradigm for evaluating surveillance effectiveness because it is not feasible, evaluators still have many useful tools at their disposal to evaluate surveillance. Using a “hard to evaluate” evaluation paradigm, in contrast to the paradigm used in the black box “easy to evaluate” situation,<sup>9</sup> involves delving into the black box and determining which elements of the results chain we can make some useful statements about. This is a less grand task than an evaluation of overall high level effectiveness, but in the case of IMF surveillance it is much more realistic and hence likely to provide useful information.

Instead of trying to assess impact on high level final outcomes where this not possible, such an approach adopts a four-part strategy:

- assessing the logical connection between the intermediate and final outcomes of the activity;
- monitoring whether progress **toward** final outcomes is being achieved (without trying to attribute the results to the specific agency being evaluated);
- asking as many strategic evaluation questions as possible about how activity might be improved including (but not limited to) attributional questions at the highest feasible and affordable level; and

---

<sup>8</sup> It is important to avoid the evaluation fallacy that just because it is not technically feasible to establish an attributable causal link between activity and high level outcomes that there is no such link. Whether or not attribution can be established within feasibility and cost constraints is a separate issue from the question of whether or not there is in reality a causal link between an intervention and high level outcomes.

<sup>9</sup> This is “easy to evaluate” in a conceptual sense and in no way seeks to minimize the effort that has to be put into actually executing black box type evaluation designs.

- providing information on whether the activity is meeting performance indicators that are fully attributable to it (which may often be at a relatively low level of the outcomes hierarchy).

In implementing this strategy, technical and practical feasibility considerations as well as evaluation cost become the central constraints that drive decision making about optimal evaluation resource allocation. While it is easy to encourage evaluators to attempt to jump into the black box and start making statements, there is ample potential for wasting resources on unhelpful evaluations, particularly when the subject of the evaluation is complex. An evaluation strategy is thus needed in order to ensure the efficient use of evaluation resources.

While the aim of the black box evaluation paradigm is *just to establish with some certainty whether the activity has an actual attributable effect on final outcomes*, the “hard-to-evaluate” evaluation paradigm has a different aim. This is *to optimize evaluation spending (within feasibility and cost constraints) so as to inform strategic direction, including attempting to establish attribution at the highest feasible and affordable level, while recognizing that there is always likely to be uncertainty about the attributable contribution of the overall activity to final outcomes*.

Developing an evaluation strategy now becomes all about developing an understanding of the feasibility, costs and benefits of answering different evaluation questions. Once these have been established, choices can be made to select a set of priority evaluation questions.

Priority evaluation questions are those that will provide stakeholders with the most useful information for the purpose of making future strategic decisions. But the simple paradigm’s question ‘did the overall implementation of the activity have a definite attributable effect on final high level outcomes?’ is well and truly discarded because it has been determined to be impossible to answer and it is therefore a waste of resources to attempt it. This can be an uncomfortable message to give policy-makers, who naturally want to know what exactly is being achieved, in terms of final outcomes, for the \$240 million spent on surveillance. But that question simply cannot be answered. As is so often the case in policymaking, deciding the strategic direction in which to take IMF surveillance will of necessity involve decision making under uncertainty as to the amount of overall resources worth dedicating to the activity.

### III. THE REMLOGIC APPROACH TO EVALUATION

This paper proposes the use of an evaluation planning methodology—REMLogic<sup>10</sup>—to develop an evaluation strategy which is also linked to ongoing planning and monitoring of

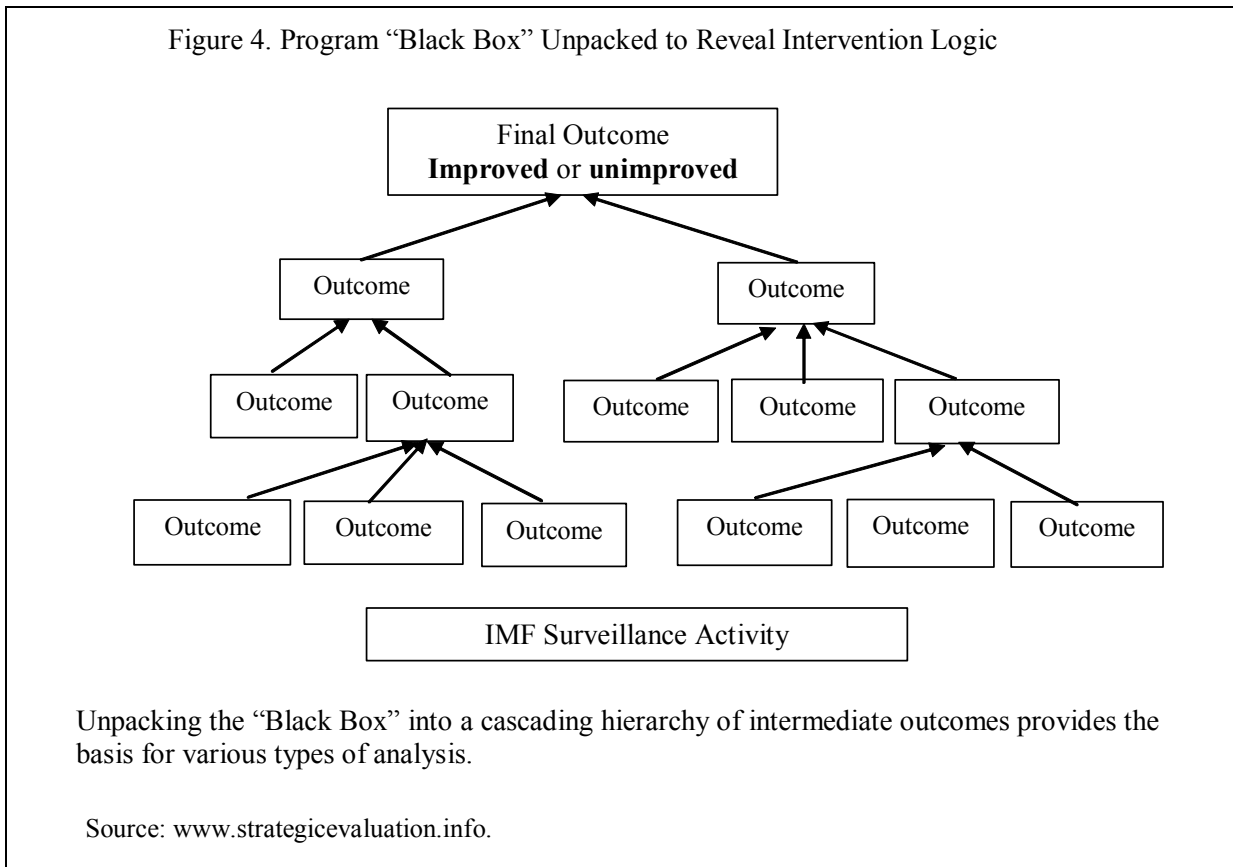
---

<sup>10</sup> Research, Evaluation, Monitoring Intervention Logic Outcomes (REMLogic) methodology was specifically developed for evaluating “hard to evaluate” activities within the Strategic Evaluation Approach to evaluation. More information on the Strategic Evaluation Approach and specific methodologies can be found at [www.strategicevaluation.info](http://www.strategicevaluation.info).

IMF surveillance. The proposed approach involves systematically working through a set of twelve questions, listed in Appendix 3 and discussed in detail here. We give an illustrative example of what might be involved at each step.

**Step 1: How is IMF surveillance believed to work?**

To answer this, the “black box” linking IMF surveillance to its final outcomes is unpacked. A logical pathway needs to be identified linking each intermediate step or outcome through which IMF surveillance activity contributes right through to its final outcomes. This is referred to as an *outcomes hierarchy*,<sup>11</sup> as generically illustrated in Figure 4. A full outcomes hierarchy for surveillance is contained in Appendix 1.

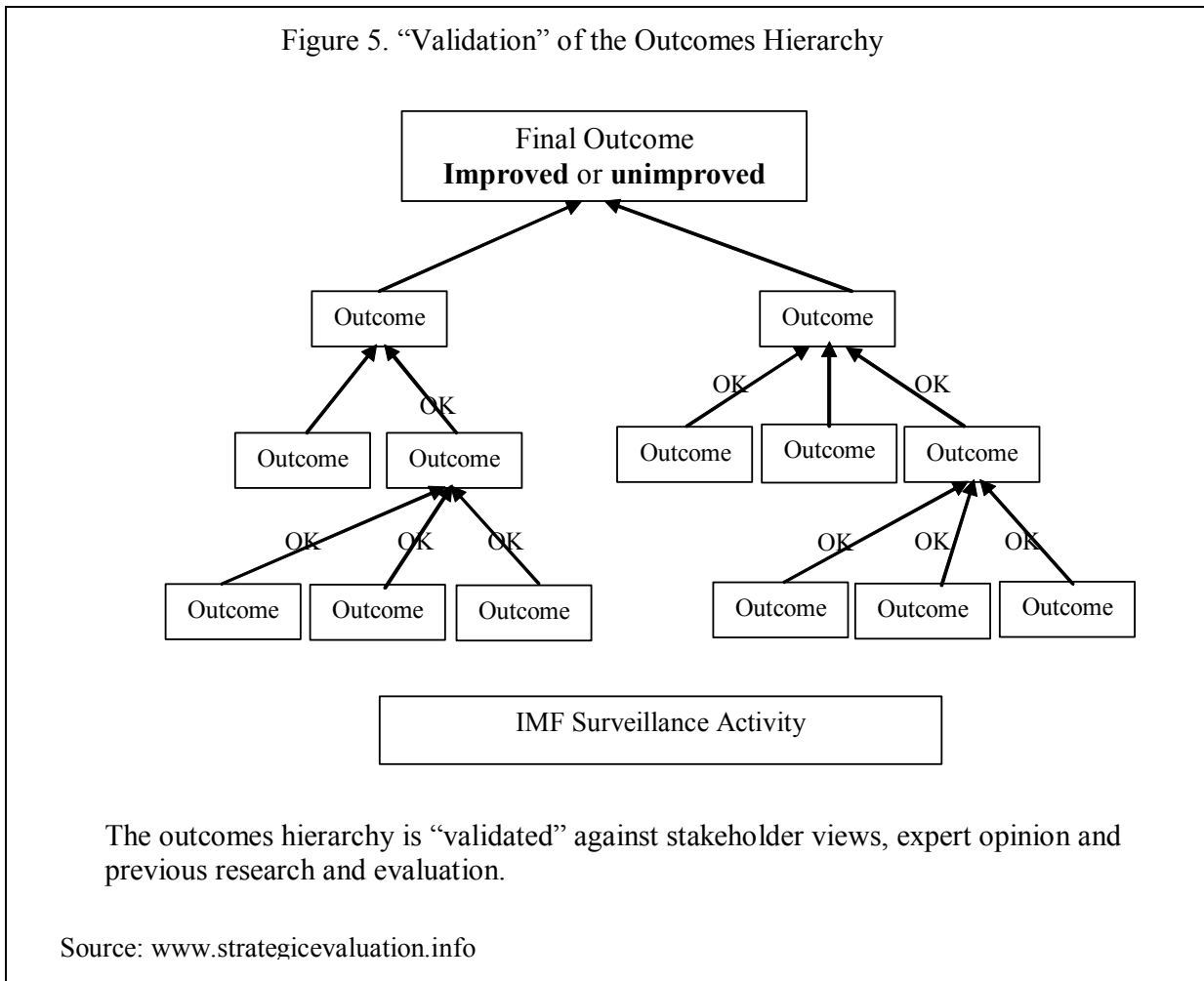


<sup>11</sup> Another accepted name for this in the evaluation literature is an *intervention logic*. We avoid this term because in the context of IMF surveillance it can be confused with the concept of making a decision regarding whether to intervene in a particular situation. The methodology used to develop the outcomes hierarchy is set out in Duignan, P. (2004), “Intervention logic: How to build outcomes hierarchy diagrams using the OH Diagramming Approach,” available at [www.strategicevaluation.info/se/documents/124pdf.html](http://www.strategicevaluation.info/se/documents/124pdf.html).

For example, part of the outcomes hierarchy for IMF surveillance may be that the outcome of *global systemic stability* is contributed to by *orderly market behavior*; which is contributed to by *appropriate individual country policies*; which is contributed to by *alignment of views of country authorities, other stakeholders and the IMF*; which is contributed to by *effective communication and publicity of IMF surveillance assessments and appropriate IMF policy advice*; which is contributed to by *accurate IMF surveillance assessments*.

**Step 2: Is the way surveillance is believed to work consistent with stakeholders' views, expert opinion and past evaluation findings?**

To answer this, the outcomes hierarchy which has been produced should be validated against the views of stakeholders about how surveillance works as well as previous research and evaluation as illustrated in Figure 5.



It is possible that different stakeholders, including policymakers in different branches of government or across different types of economies, market players or academics have quite different perceptions of how surveillance works and the ends that different components of surveillance activity are aimed at achieving.<sup>12</sup> In regard to validating the IMF surveillance outcomes hierarchy against stakeholder views and expert opinion, stakeholders or experts may suggest improvements in the outcomes hierarchy. Outcomes hierarchies can be struck at various levels, both at the overview level of how surveillance operates as is being done in this paper, and at more specific levels, for instance, the mechanism linking country actions and country outcomes. Detailing such specific outcomes hierarchies/logics can be subprojects within an overall evaluation.

For example, first, the overall outcomes hierarchy of surveillance could be sent out for peer review from other institutions involved in similar work and from a selection of specific country stakeholders. Second, a subproject within the evaluation of surveillance could be for a group of specialists to detail their view of the operating mechanisms linking IMF advice to country outcomes (drawing on previous analysis and documentation on this topic) and have them set this out as an outcomes hierarchy or similar type of model. This then could be the subject of rigorous peer critique.

### **Step 3: What routine monitoring can there be of whether the intermediate and final outcomes are being achieved (i.e. strategic indicators)?**

To answer this question, each outcome in the outcomes hierarchy is examined to see if there are routine indicators that are relatively easy to measure to monitor the achievement of IMF surveillance's final and intermediate outcomes. At this stage the focus is simply on the strategic question of whether the outcomes that really matter are being achieved, or at least progress is being made in the right direction—not whether it can be definitely proved that it is IMF surveillance activity which is changing the outcomes. Therefore, such *strategic indicators* do not need to be attributable<sup>13</sup> to IMF surveillance (however, if they are attributable, so much the better). This step integrates indicator monitoring activity with evaluation planning. Figure 6 below shows strategic indicators in the first column on the right hand side. **If strategic indicators show that final and intermediate outcomes are not being achieved, there needs to be intense critique of the activity and the wider system in**

---

<sup>12</sup> To illustrate, different stakeholders might have different perspectives on what it is about the content of IMF surveillance analysis that influences the policy debate. Is it originality in analysis (i.e., saying something new)? Or is it the independence/neutrality of the analysis (even if it is not especially original)? Such differences in perspective could lead to quite different measuring rods for assessing the content of surveillance output.

<sup>13</sup> It should be stressed again that just because an outcome cannot be demonstrably linked to a particular activity says nothing about whether or not the activity is actually influencing the outcome.

**which it is operating in order to see if things could be done better because the whole purpose of the activity is not being achieved (regardless of who is accountable).**

For example, some indicators for the highest level outcome (global prosperity as defined by Article I) could be: *developments in world GDP; poverty indicators; progress toward Millennium Development Goals (MDGs)*. Similarly, various measures of inflation and global financial imbalances could be used as indicators to track global stability. IMF surveillance activity constitutes one part of member countries' strategy toward achieving good outcomes in these areas, but whether or not this happens is unlikely to be directly attributable to IMF surveillance.

**Step 4: What routine monitoring can there be which attributes intermediate outcomes to IMF surveillance activity (i.e. performance indicators)?**

To answer this question, each outcome in the outcomes hierarchy is examined to see if there are any to routinely measure *attributable indicators* that are relatively easy to measure, in order to monitor the achievement of intermediate outcomes. Since these performance indicators must be clearly attributable to IMF surveillance activity, they will tend to be at a lower level of the outcomes hierarchy than the *strategic indicators* discussed above. **These attributable indicators should be pushed as high up the intermediate outcomes hierarchy as possible**, only stopping at the point where they become impossible or too expensive to measure routinely. In a case where feasibility and cost constraints make it impossible to push very far up the intermediate outcomes hierarchy, the performance indicators will look like outputs—such as numbers of reports published, pages drafted, number of issues covered, etc.<sup>14</sup>

This step integrates output level reporting with evaluation. Since these indicators can be used to hold the activity to account for its performance, they are alternately referred to as *performance indicators, attributable indicators or accountability indicators*. Figure 6 below shows the performance indicators in the second column on the right hand side.

For example, performance indicators for IMF surveillance could be the *delivery of input into multinational policy fora*, or the *release of IMF publications into the public domain*, or *coverage of FSAP reports*, perhaps accompanied by measurement of the coverage of policy issues in the reports or the number of press conferences or statements given to the media.

**Step 5: Assess how comprehensively the strategic and performance indicators cover the outcomes hierarchy.**

---

<sup>14</sup> Depending on the way they are defined, outputs do not have to be mutually exclusive of outcome indicator measurement. Demanding mutual exclusivity creates certain technical problems. See the discussion of Outcomes Theory at [www.strategievaluation.info](http://www.strategievaluation.info).

To do this, the coverage of indicators across the outcomes hierarchy needs to be reviewed. This process allows decisions to be made about the optimal mix of indicators which should be collected. Figure 6 sets out a schematic presentation of the outcomes hierarchy, the strategic indicators and the performance indicators. There is likely to be both areas of indicator overlap and some blank areas within the outcomes hierarchy.

If it is found that there are no indicators in one part of the logic and a disproportionate number in another part, it may be appropriate to put more resources into developing some indicators in areas that are not currently well provided for. If there are a number of higher level indicators in either the strategic or performance indicator set then this may suggest some redundancy in indicators at a lower level within the set (because it is already known how higher level indicators are tracking). Such judgments will be an important part of the ongoing effort to develop better indicators to track the effectiveness of surveillance.

#### **Step 6: What important evaluation questions can be asked for each intermediate outcome or group of intermediate outcomes?**

To address this, each intermediate outcome in the outcomes hierarchy is examined to see what evaluation questions could be asked about it. In addition to questions about individual outcomes, there may be some groups of intermediate outcomes or relationships between groups of outcomes for which particular evaluation questions can be asked.

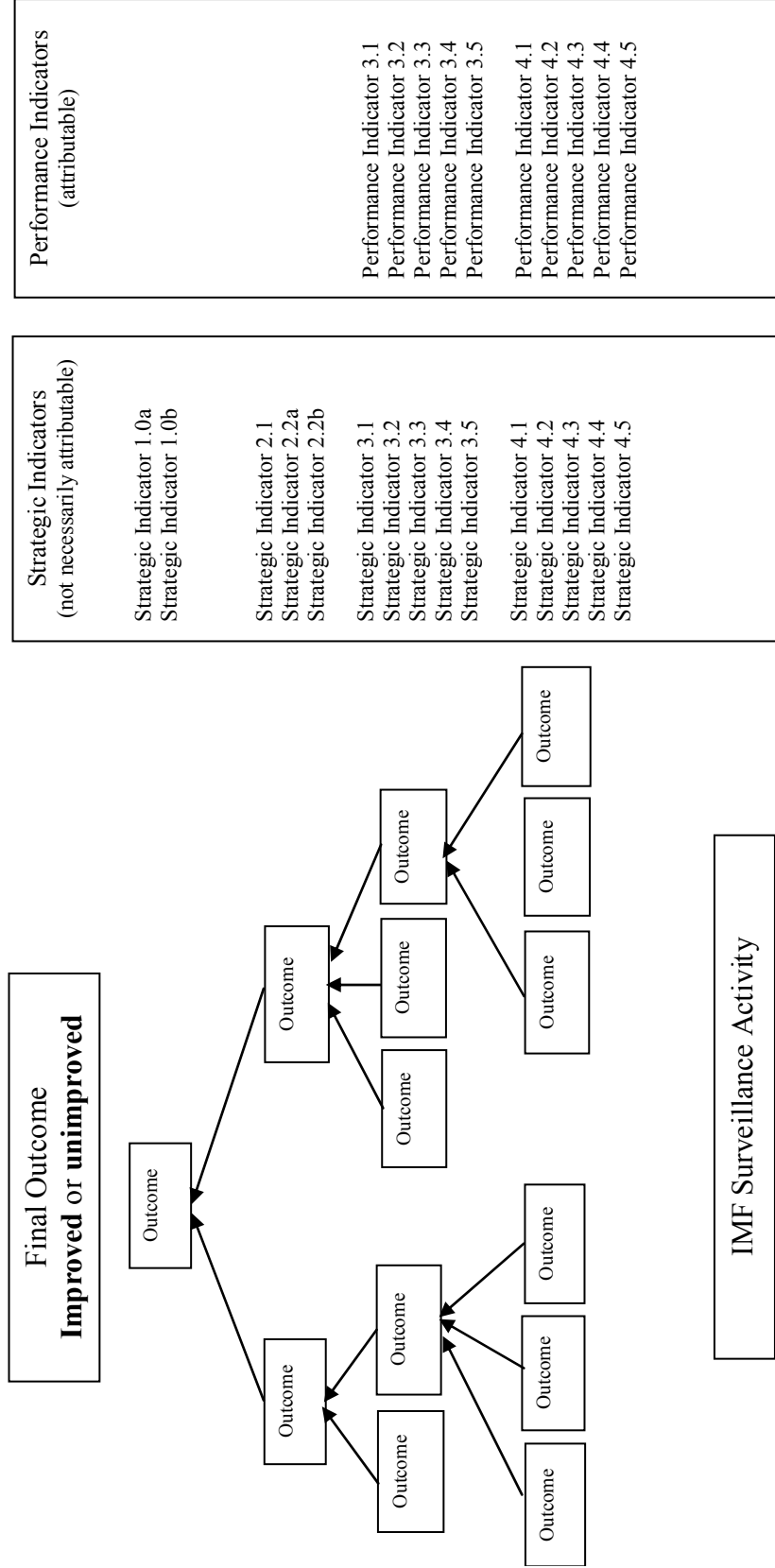
For example, in regard to the two intermediate outcomes “accurate IMF assessment” and “appropriate IMF policy advice,” evaluation questions could be “how accurate are IMF assessments” and “how appropriate is IMF advice.” In regard to the intermediate outcome “appropriate behavior by country,” an evaluation question could be “what effect does IMF surveillance have on individual country behavior and through what channels does it operate?”

#### **Step 7: What evaluation questions are technically and practically feasible to answer?**

In this step, the technical feasibility of answering each of the identified evaluation questions is examined. As discussed earlier, in many cases, for questions related to high-level and final outcomes, it may be impossible to answer some evaluation questions because of the nature of the activity and its context. However, **before any high level attributional evaluation questions are dismissed as not technically or practically feasible, it is essential that considerable thought be put into the feasibility of answering them.** These are after all the questions which, if answered, would provide the greatest yield of strategic information. The technical and practical feasibility of answering each evaluation question could be rated into broad ranges such as: currently impossible, high difficulty, medium difficulty and low difficulty.

For example, the evaluation question “how accurate are IMF assessments” is probably low difficulty and “how appropriate is IMF advice” is probably medium difficulty. The evaluation question “what has been the impact of FSAP exercises” on measures of financial stability is probably not possible to answer at this time.

Figure 6. Outcome Hierarchy, Strategic and Attributable Performance Indicators



Source: [www.strategievaluation.info](http://www.strategievaluation.info).



**Step 8: What is the likely cost of answering technically and practically feasible evaluation questions?**

The cost of answering particular evaluation questions is estimated into the broad ranges: high, medium, low.

For example, the evaluation questions: “what is the effect of transparency (such as publication or non-publication of IMF surveillance reports) on communicating assessments to country authorities” and “is there any loss of candor from transparency” probably medium to low in terms of cost to answer.

**Step 9: What are priority evaluation questions within available evaluation resources?**

The feasibility and cost of all of the potential evaluation questions are balanced against expected benefits of obtaining answers, in order to identify those which are the highest priority in terms of informing future IMF strategy development for surveillance. In prioritizing evaluation questions, the information needs of both the IMF and of other stakeholders should be taken into account. Step 9 ensures that evaluation spending is tightly targeted on answering priority evaluation questions for future strategy development.

At the level of determining its broad work program, the IEO has already established and presented to the IMF Board a set of criteria for choice of evaluation projects. This considers such issues as relevancy to a wide IMF membership, potential for enhancing effectiveness vis a vis the IMF mandate, the stakes in terms of financial and reputational risk for the institution of the issue being evaluated, and the issue’s ripeness for evaluation. Determining priority evaluation questions within broad areas such as surveillance will require a similar exercise of balancing costs against benefits. The power of the REMLogic methodology lies in the fact that it forces explicit, transparent and peer-reviewable decision making about exactly which evaluation questions are, and are not, being attempted. This is in contrast to evaluation strategies which claim to be measuring “the effectiveness of IMF surveillance” without clearly identifying exactly which evaluation questions they will be answering by mapping them onto the underlying outcomes hierarchy. In order to generate stakeholder confidence that an institution is not attempting to avoid high level attributional questions which are actually technically feasible and affordable, the priorities determined through any REMLogic exercise should be subject to rigorous stakeholder review. The process of circulating for public comment the draft issues papers/terms of reference for each IEO evaluation project provides a good vehicle for such a review.

To take just one example from within the outcomes hierarchy, the evaluation question “does IMF multilateral surveillance result in peer pressure from other countries” is probably technically and practically feasible to answer via survey or interview methods, and if costs are deemed reasonable, it may be determined to be a priority evaluation question.

### **Step 10: How can the evaluation questions be grouped into projects and phased?**

Evaluation questions that use similar methodologies, or that are closely related in terms of their implications for the institution, can be grouped into evaluation projects, and the phasing of the rollout of these evaluation projects can be decided. Evaluation projects can then be worked into an Evaluation Issues Paper/Terms of Reference and made available to the stakeholders at an early stage by posting on the IEO website and circulating to the Board, IMF staff and selected outside experts for comments before finalization. These papers spell out the objectives, logic and process of a specific evaluation project.

### **Step 11: Collecting lessons from the priority evaluation projects**

Evaluation findings are mapped back onto the outcomes hierarchy as they flow in. If the outcomes hierarchy is used in strategic planning, Step 11 ensures that evaluation results are directly linked back into organizational learning and strategy rather than passing under the radar screen as scattered reports of individual evaluations. This step requires an underlying organizational knowledge and familiarity with management strategy.

For example, the results from an evaluation project answering the question “how effectively are results from the FSAP exercises incorporated into IMF surveillance” would be fed back directly into strategic planning when improvement in achieving the relevant intermediate outcomes is being considered.

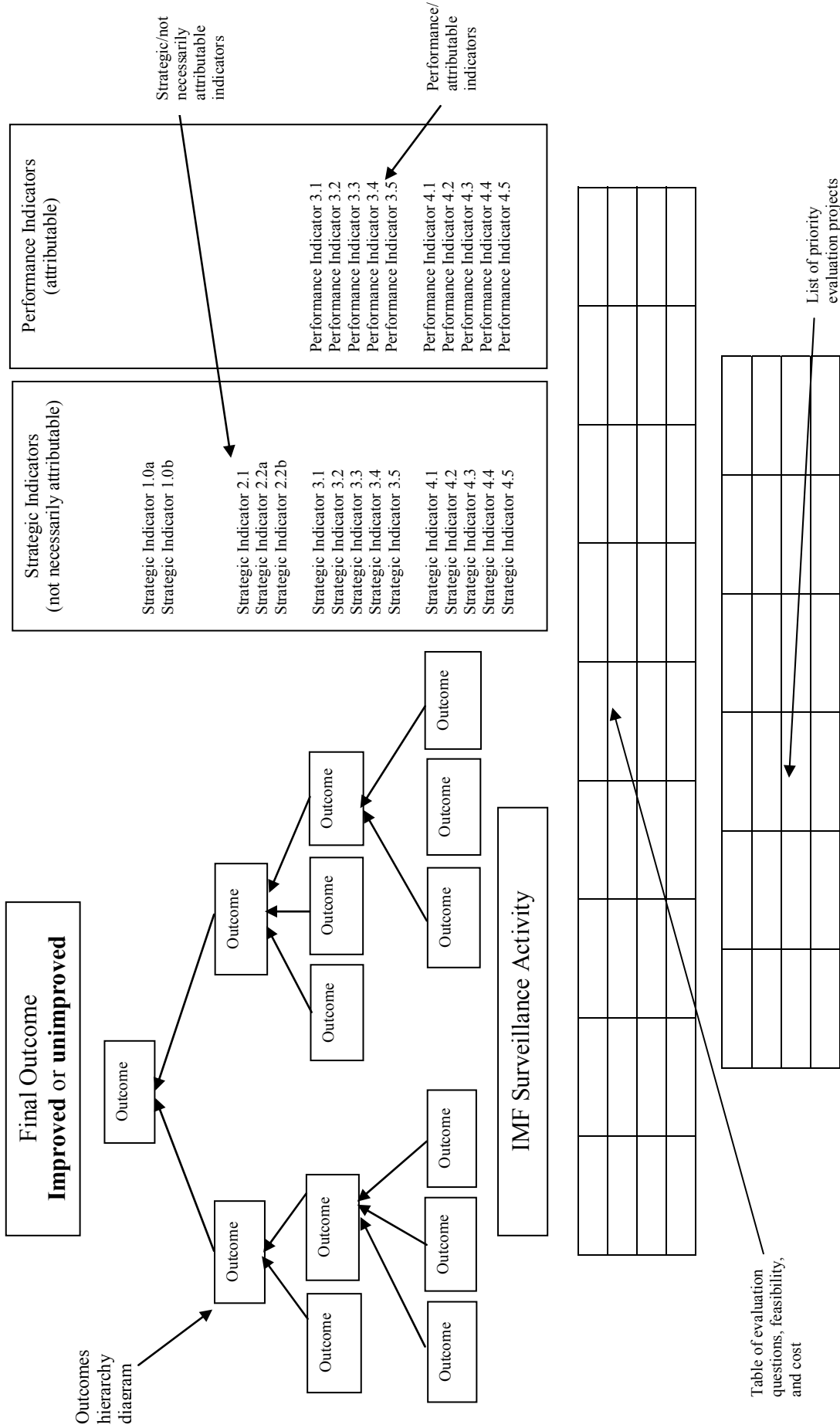
### **Step 12: Determining the next priority evaluation questions**

The whole set of questions is asked again at regular intervals and a new set of evaluation priority projects is developed as the evaluation strategy evolves.

What emerges from systematically answering the questions set out above is a *REMLogic Structure* for an organization or area of activity. Such a structure is illustrated in Figure 7 and consists of:

- a diagram of the outcomes hierarchy;
- two sets of indicators (not-necessarily attributable strategic indicators and clearly attributable performance indicators);
- a table of evaluation questions (with feasibility, cost and priority identified); and
- a list of priority evaluation projects.

Figure 7. Outcome Hierarchy, Strategic and Attributable Performance Indicators



Performance Indicators (attributable)

Strategic Indicators (not necessarily attributable)

Strategic/not necessarily attributable indicators

Performance/attributable indicators

- Performance Indicator 3.1
- Performance Indicator 3.2
- Performance Indicator 3.3
- Performance Indicator 3.4
- Performance Indicator 3.5
- Performance Indicator 4.1
- Performance Indicator 4.2
- Performance Indicator 4.3
- Performance Indicator 4.4
- Performance Indicator 4.5

- Strategic Indicator 1.0a
- Strategic Indicator 1.0b
- Strategic Indicator 2.1
- Strategic Indicator 2.2a
- Strategic Indicator 2.2b
- Strategic Indicator 3.1
- Strategic Indicator 3.2
- Strategic Indicator 3.3
- Strategic Indicator 3.4
- Strategic Indicator 3.5
- Strategic Indicator 4.1
- Strategic Indicator 4.2
- Strategic Indicator 4.3
- Strategic Indicator 4.4
- Strategic Indicator 4.5

Table of evaluation questions, feasibility, and cost

List of priority evaluation projects

#### IV. APPLYING THE REMLOGIC APPROACH TO SURVEILLANCE EVALUATION

Since IMF surveillance activity is deeply embedded within the operations of the Fund, it is important for any evaluation strategy to take into account and make use of other organizational activities, particularly review and monitoring processes, performance indicator measurement strategic planning. This is both more efficient because it avoids duplication, and, if done well, it links evaluation with other organizational activity in integrated support of a *learning organization*.

Other organizational activity related to IMF surveillance includes, *inter alia*, the regular internal biennial reviews of surveillance; an ongoing review of the IMF's work on standards and codes; and work that is currently underway in PDR on developing routine indicators for surveillance activity. In planning for the IEO's evaluations of surveillance, it is essential to take this other organizational activity into account, so that the IEO's work complements rather than duplicates, what is already being done internally.

The design of the REMLogic methodology supports this type of integration. For instance, in contrast to many methods of evaluation planning, it includes two types of indicator measurement. This is fully consistent with performance measurement approaches, and takes them further by ensuring that both strategic indicators and performance indicators are collected and differentiated where necessarily—strategic indicators for periodic strategic purposes and performance indicators for regular accountability and efficiency improvement purposes. The REMLogic methodology also allows a place for appropriate feedback to other internal organizational review/monitoring processes and strategic planning, while at the same time accommodating the role of independent external evaluation.

At the operational level, the REMLogic methodology has been applied to IMF surveillance activity and a draft evaluation structure (outcomes hierarchy diagram, a set of strategic indicators some of which are not attributable, a set of performance indicators that are by definition all attributable, a table of evaluation questions and a list of evaluation projects) has been developed.

Appendix 1 sets out an outcomes hierarchy diagram for IMF activity. In the diagram, all the possible linkages between intermediate outcomes are not spelled out, in the interests of keeping the diagram manageable; rather, the shape of the boxes in which the intermediate outcomes are placed is used to indicate the way the intermediate outcomes interrelate.

Note that some elements of the outcomes hierarchy are unrelated to surveillance. They are, however, highly related to intermediate and final outcomes in the hierarchy, and as such their exclusion would be inappropriate when mapping the channels through which IMF inputs and outputs flow into desired outcomes.<sup>15</sup> It should be noted that this simplified outcomes

---

<sup>15</sup> In the REMLogic approach the question of which attributional questions will be asked is not dealt with by only allowing certain outcomes to appear in the outcomes hierarchy, but

(continued...)

hierarchy and the more complex working hierarchy within the REMLogic structure on which it is based is only the first iteration of this outcomes hierarchy. Ongoing critique, validation and improvement of this hierarchy is a central part of the process.

### **A. Making the REMLogic Approach Operational: A Concrete Example**

The process by which the IEO has approached the evaluation of multilateral surveillance provides insight into how this type of framework can be operationalized. The results of this process are reflected in the draft issues paper for the evaluation project, which is available on the IEO website at [www.imf.org/ieo](http://www.imf.org/ieo).

Step 1 of developing the evaluation strategy involved arriving at a consensus on how multilateral surveillance is supposed to work. This involved unpacking the “black box” linking activity to final outcomes and drawing up an outcomes hierarchy.

An important characteristic of surveillance is that it is interwoven into virtually all of the IMF’s work and outputs, so that it has stakeholders both internally and externally. Moreover, there did not seem to be a clear operational distinction between multilateral surveillance and other IMF surveillance. As a consequence, there were from the outset differences of opinion on how best to unbundle surveillance from other IMF activities, and multilateral surveillance from the rest of surveillance.

As a practical matter, drawing an outcomes hierarchy needed to start from the top and work down, with the highest level final outcome (i.e. IMF Article I objectives) placed in a box numbered 1.1. Underneath this box came a level of high-level intermediate outcomes (labeled 2.x) which collectively satisfy what is needed to achieve 1.1. Below these came progressively lower level intermediate outcomes, all in suitably numbered boxes.

Step 2 in the process involves taking our internal consensus outcomes hierarchy mapping to a wider audience of stakeholders, both internal and external, to validate the various linkages that have been identified. As a practical matter, there are limits to how many people can usefully be incorporated in the initial stages of drawing up a hierarchy. Nevertheless, restricting the number of participants risks missing important views.

In this case, the issues paper did not actually include the outcomes hierarchy mapping contained in Appendix 1; many stakeholders are only be interested in one particular link and so find the rest of the mapping to be largely irrelevant to their needs. Exactly how the logic is communicated to different groups of stakeholders will depend on the project. Operationally in the case of multilateral surveillance, comments from stakeholders on the issues paper could be matched against the linkages that have been identified, and to the extent that new

---

within the specific evaluation questions being asked. This prevents the outcomes hierarchy being distorted by the issue of whether attribution is feasible and affordable. For more on this issue see the discussion of Outcomes Theory on [www.strategicevaluation.info](http://www.strategicevaluation.info).

linkages emerge, the outcomes hierarchy mapping might be modified. In this way, it is easier to organize various links into meaningful evaluation questions that address key issues without overemphasizing some links by asking essentially the same question in multiple ways, and ignoring others.

Likewise, past evaluations of surveillance, particularly the three most recent biennial reviews of surveillance and the associated Board discussions, were studied carefully for insights that have already been internalized as regards to some linkages, and for identification of linkages that have yet to be evaluated more formally.

The solicitation of comments illustrates the iterative procedure involved in settling on a workable outcomes hierarchy mapping.<sup>16</sup> Because the mapping may change in some material ways during this process, it is essential that a consistent numbering system exists which links outcomes boxes to associated indicators, which are listed in separate tables (more below).

Steps 3 and 4 involve asking what indicators are available regarding the achievement of intermediate and final outcomes. This step is completely compatible with what has already been suggested internally within the IMF regarding performance indicators, but it takes the process one step further. In particular, a distinction is made between performance indicators, which are directly attributable to IMF activity, and strategic indicators, which are not necessarily attributable (such as correlation between IMF advice and adopted policy, or absence of financial instability). It is important to present both lists of indicators simultaneously. Presenting only the list of strategic indicators will frustrate the thoughtful reader, as they do not necessarily provide any information about attribution. Likewise, presenting only performance indicators will, in cases where it is difficult to attribute, ignore that part of the outcomes hierarchy which is the most interesting for decision makers, i.e., the part that responds to the question of whether progress is being made with regard to the higher-level strategic objectives of surveillance.

Steps 5, 6, 7 and 8 involved studying the extent to which strategic and performance indicators cover the entire outcomes hierarchy, and making a table of possible evaluation questions associated with each box and link in the outcomes hierarchy, including the likely cost of answering these questions. Two observations stood out in the use of the REMLogic method in regard to multilateral surveillance. The first was that it made little sense to try to identify indicators and associated questions for every box in the outcomes hierarchy; many were simply not particularly relevant to the evaluation project at hand. As a consequence, the resulting table of evaluation questions had many blank spaces in it. A complete table might have a useful database-type function in quickly sorting out how to evaluate any of several

---

<sup>16</sup> While outcomes hierarchies do need to be stabilized for evaluation planning, they never become permanent. At some point new findings or analysis will lead to some changes in the causal pathways in an outcomes hierarchy, reflecting an evolving world and institutions.

surveillance-related questions, but constructing such a database in the context of our exercise seemed superfluous.

The second observation was that while it was possible to find various indicators for most boxes, and even at the highest levels of strategic outcomes the indicators were surprisingly easy to collect, it was also true that performance indicators were almost exclusively concentrated at the lower levels of the hierarchy and strategic indicators at the higher ones as is often the case. The challenge is to try to push performance indicators as far up the outcomes hierarchy as possible subject to feasibility, cost and priority considerations.

Step 9 and 10 involved prioritizing evaluation questions, and grouping them into subprojects. In a sense, this step was partly done at a much earlier stage when we eliminated certain branches of the outcomes hierarchy from the evaluation at hand. What was left was a number of identifiable logical chains of intermediate outcomes with associated indicators and evaluation questions. These were described in the draft issues paper mentioned in Step 2 above, and a review of prioritization across subquestions will follow after a consultation process with key stakeholders of surveillance helps determine where the most important issues lie and the tradeoffs involved in evaluating one area versus another.

Steps 11 and 12 are follow-up steps to be implemented only after the evaluation has been completed.

## **B. Specific Evaluation Methods and Designs for Collecting Evidence**

There is a wide range of evaluation methods and designs available to collect evaluation evidence. Evaluation methods are specific methods used to collect information in an evaluation project, for example, *surveys* or *interviews*. Evaluation designs are the way in which an evaluation project is structured, for example, as an *experiment* or a *case study*.<sup>17</sup>

While this paper is not intended as a detailed guide to specific methods (and methodological pitfalls), the most common types of methods likely to be used in evaluations of IMF surveillance would include:

- *Document thematic analysis* (sometimes referred to as *desk reviews*). Analysis of documents in order to extract themes in regard to the topic being evaluated. This can include all types of documents such as formal reports, minutes of meetings, memos and print media reports. Where the documents are research reports this is referred to as a literature review. The analysis can range from an extraction of general themes to a tight and specific detailed analysis (e.g., a rating of how well an IMF surveillance report addresses a specific issue). If qualitative ratings are used, an element of

---

<sup>17</sup> This discussion is drawn from Duignan, P. *Introduction to Strategic Evaluation: Evaluation Approaches, Purposes, Methods and Designs*.  
[www.strategievaluation.info/se/documents/104f.html](http://www.strategievaluation.info/se/documents/104f.html).

subjective judgment is inevitable but the risks can be reduced by documenting as explicitly as possible, in advance, what the reviewer would expect to see to warrant specific ratings and by using more than one reviewer.

- *Written surveys.* Surveys can be undertaken with groups of stakeholders who are the users of, or knowledgeable about, aspects of the effect of IMF surveillance activities. In such surveys, respondents can be questioned on how they view both the quality and the results of surveillance. The advantage of respondents' judgments of this type is that they can provide responses which combine multiple factors into an overall summary evaluative judgment. The disadvantage is that such respondents may not be fully knowledgeable or may not, because of their close involvement in using IMF surveillance products, be able to provide a sufficiently disinterested view on IMF surveillance. Written surveys also have the disadvantage of potentially low response rates. In undertaking written surveys on IMF surveillance there needs to be clarity about whether an individual response or a consensus country response is being sought for the evaluation.
- *Telephone or face-to-face interviews.* Stakeholders can also be interviewed either by telephone or face-to-face. The issue of respondents' judgments discussed above in regard to written surveys also applies to telephone and face-to-face interviews. However, both telephone and face-to-face interviews have the great advantage of allowing for interaction; such interaction lets the interviewer pursue details of points respondents make both in the interview itself and also to seek subsequent interviewees' perspectives. However, undertaking interviews is more expensive than written surveys, and therefore fewer respondents will be able to be contacted for a similar cost. In most settings, telephone or face-to-face interviews achieve a higher response rate than written surveys. The issue of whether an individual or a consensus country response arises in interviews is the same as in written surveys. If a country consensus perspective is being sought in an interview, this should be clarified well in advance of the interview taking place so that the interviewee has time to collect a consensus view.
- *Stakeholder focus groups and group interviews.* Group interviews are used as a cost-effective way of interviewing more than one stakeholder at the same time. There is nonetheless a trade-off between group interviews and individual interviews in terms of the amount of time each stakeholder has to respond. Focus groups are group interviews where there is an interest in having participants discuss an issue *amongst themselves* in order to provide a richer perspective on the topic than can be obtained by either individual interviews or group interviews where respondents spend all the time just answering the questions as individuals.
- *Statistical and econometric analysis.* Statistical and econometric analysis is any quantitative data analysis. The data being analyzed can be routine statistics collected by the IMF or other bodies or the results of data collection within specific evaluation projects.



*Evaluation designs* are the way in which evaluation methods are used and in some cases combined in order to draw out evaluation conclusions. Evaluation designs can incorporate any of the methods listed above. Possible evaluation designs include:

- *Experimental or quasi-experimental<sup>18</sup> comparison of changes in approach.* While an experimental “black box” approach could not be used to assess the overall effect of IMF surveillance activity on final outcomes, a quasi-experimental method may be applicable at lower levels in the outcomes hierarchy. It should be considered in any instance where one group can be treated in a different way from another group and the resultant outcomes measured. For instance, if a novel method of doing a particular type of surveillance was being considered by the IMF (arguably the case as regards regional surveillance conducted in some area departments or with “pilot projects” on approaches to financial sector surveillance), it may be possible to use the new method for the surveillance of one group of countries, use the current surveillance method on another group and then compare the outcomes in terms of the quality of surveillance reports produced and/or the reactions of stakeholders to the surveillance activity.
- *Regression discontinuity designs.* Where entities such as countries can be rated in terms of an outcome variable and where it is ethical and practical to only provide activity such as some novel aspect of surveillance (“the treatment”) to a sub-set of countries, it might be theoretically possible to use a regression discontinuity design. This is where the treatment is only provided to those countries below a cut-off point on the outcome variable. Once sufficient time has passed for the treatment to have had an effect, the new data on the outcome variable is examined for each country and if the novel approach has had an effect there should be a clear discontinuity at the cut-off point on a graph of outcomes for all countries.
- *Time series designs.* Time series designs are where trends in data are examined over time to determine whether they are associated with changes in treatments to entities such as countries. These require clearly identified changes in treatment and may or may not be interpretable depending on the specific situation. For example, various aspects of surveillance in countries with and without IMF-supported programs could be amenable to such analysis.
- *Single country case studies.* Single country case studies look in detail at what happened in the case of IMF surveillance of a single country. Using a case study methodology they examine in detail exactly what happened and the role of IMF surveillance in the case of the single country. Once this work has been done, conclusions may be able to be drawn as to how the role of IMF surveillance could be

---

<sup>18</sup> *Quasi-experiments* are designs which attempt to follow an experimental approach but in which various compromises have been made because of the reality of program implementation.

improved in dealing with similar countries in the future. Past IEO evaluations of the IMF's role in three capital account crisis cases (Brazil, Korea and Indonesia) and in Argentina have contained such assessments of surveillance.

- *Event case studies.* Event case studies track the course and ramifications of a particular incident which has affected the world economy. Using a case study methodology they examine in detail exactly what happened and the role of IMF surveillance in the way the event unfolded. Once this work has been done, conclusions may be able to be drawn as to how the role of IMF surveillance could be improved in dealing with similar events in the future. Such an approach could be well-suited to assessing how well different aspects of IMF surveillance (bilateral, multilateral) have dealt with global interlinkages arising from specific events.
- *Correlational and cross-sectional comparative studies.* These designs compare one variable with one or more other variables. They can be used, for instance, to compare the results of IMF surveillance forecasting with what occurred in reality and with similar forecasts made by other agencies. The implications of the results from such analyses need to be interpreted carefully because of the *prediction-prevention effect* and the *self-fulfilling prophecy* effect. The prediction-prevention effect is where negative predictions about the future encourage stakeholders to take actions which in turn prevent the predicted negative events from taking place. The self-fulfilling prophecy effect is where a prediction, e.g., of a crisis situation, results in stakeholders acting in ways that exacerbate the problem which has been predicted. These two effects make it difficult to interpret the findings from correlational studies in a number of cases.
- *Exhaustive causal identification and alternative explanation elimination design.* These designs attempt to identify all of the possible causes for an outcome, collect evidence for and against each of them and through a process of elimination identify one or more causes for an outcome. This is in contrast to some pseudo-outcome studies which simply collect evidence consistent with attribution of changes in outcomes to one treatment without examining whether the evidence is not also consistent with attributing changes in outcomes to alternative factors. Due to the large number of factors affecting the high-level outcomes in the case of IMF surveillance it seems unlikely that this methodology could be applied at a very high-level. However, it may be able to be used at lower levels of the IMF surveillance outcomes hierarchy.

## V. CONCLUSIONS

Surveillance is rightly one of the most frequently evaluated activities of the IMF. While most evaluation activity is internal and geared toward incrementally improving the practice of surveillance, there is also a role for occasional external evaluation with a view to providing transparency and accountability. An evaluation strategy for IMF surveillance needs to find an appropriate balance between the two elements.

Consideration should also be given to what either evaluation is reasonably able to deliver. Because of the nature of IMF surveillance, and in particular the absence of any counterfactual, it is unrealistic to expect much in the way of an assessment of the overall effectiveness of surveillance at the very highest outcomes level. Nevertheless, increasing effectiveness of surveillance is of major importance, and well-designed evaluations should be able to serve a very useful purpose in improving the operation of surveillance.

The design of an evaluation is a non-trivial problem. For both transparency reasons as well as to avoid wasting time and effort on unhelpful evaluation activity, careful thought needs to go into initially drawing up a reasonably detailed hierarchy of outcomes or a results chain which most stakeholders will regard as broadly correct, after which decisions need to be taken with regards to which nodes of the outcomes hierarchy would be the most critical for decision makers to have more information on. Consistent pressure needs to be maintained on attempting to establish attribution at the highest possible level to prevent any institutional attempt at avoiding the most powerful and useful questions external stakeholders want answered. However, pushing such questions beyond what is technically feasible does a disservice to all concerned by producing useless and obfuscating “pseudo-outcome evaluations.” Once the critical nodes have been identified, there are a number of available techniques that can be employed to answer key evaluation questions, which each individually or in combination have particular pros and cons.



**EXAMPLES OF POTENTIAL EVALUATION QUESTIONS**

As an illustration, the table below provides some selected examples from the Table of Evaluation Questions within the draft REML logic structure which has been developed for surveillance so far. The (intermediate) outcome is listed at the far left, followed by strategic and performance indicators in columns 2 and 3 respectively. The remaining columns list previous evaluation findings, potential evaluation questions, assessments of technical and practical feasibility, estimated cost, and priority. It should be noted that this table is for illustrative purposes only and the ratings regarding the evaluation questions may well change following further consideration.

<b>Outcome</b>	<b>Potential strategic indicators (may not be attributable)</b>	<b>Potential performance indicators (clearly attributable to IMF activity)</b>	<b>Previous evaluation findings</b>	<b>Potential evaluation questions</b>	<b>Technical and practical feasibility of answering the question</b>	<b>Estimated cost</b>	<b>Priority</b>
Article I objectives (economic prosperity)	Development in world GDP; poverty indicators; progress towards MDGs; distribution of regional GDP growth	Currently no indicator (Indicator development: not currently being attempted)	Previous evaluations findings at lower level in outcomes hierarchy	What is the effect of IMF surveillance on global prosperity?	Likely to be impossible	N/A	N/A
Appropriate behavior by individual country	Various indicators of macroeconomic outcomes and policies for the individual country Market risk premia on sovereign debt	Currently no indicator (Indicator development: IMF Board has called for surveillance reports to discuss authorities' follow up to previous surveillance policy advice	Some information from previous event studies	Is appropriate behavior by individual countries influenced by IMF surveillance?	High difficulty	Medium to high	High

**Examples of Potential Evaluation Questions (continued)**

<b>Outcome</b>	<b>Potential strategic indicator (may not be attributable)</b>	<b>Potential performance indicators (clearly attributable to IMF activity)</b>	<b>Previous evaluation findings</b>	<b>Potential evaluation questions</b>	<b>Technical and practical feasibility of answering the question</b>	<b>Estimated cost</b>	<b>Priority</b>
Appropriate IMF policy advice	None	Currently no indicator (But Board has established several monitorable benchmarks with regard to priorities for coverage of surveillance)	Previous evaluation findings regarding appropriateness of policy advice	How appropriate is IMF policy advice?	Medium difficulty	Medium	High
Accurate IMF assessments	Same as attributable performance indicator	Regular comparisons of IMF World Economic Outlook forecasts with other agency forecasts and outcomes	Previous evaluation findings regarding assessments	How accurate are IMF assessments?	Medium difficulty	Low to medium	High
Effective, accurate and appropriate communication of IMF assessments and policy advice	Same as attributable performance indicator	Monitor of media coverage of sample of assessments (Indicator does not cover appropriateness)	Previous evaluation findings regarding transparency	What is the effect of transparency (i.e. publication or non-publication of IMF surveillance reports) on communicating assessments to country authorities?	Low difficulty	Low	Medium

**Examples of Potential Evaluation Questions (concluded)**

<b>Outcome</b>	<b>Potential strategic indicator (may not be attributable)</b>	<b>Potential performance indicators (clearly attributable to IMF activity)</b>	<b>Previous evaluation findings</b>	<b>Potential evaluation questions</b>	<b>Technical and practical feasibility of answering the question</b>	<b>Estimated cost</b>	<b>Priority</b>
Effective integration of analysis across departments	None at present	None at present	Previous evaluation suggest inadequate integration of bilateral and multilateral surveillance	Has integration of bilateral and multilateral surveillance improved and what organizational aspects influence it?	Medium	Medium	High
Appropriate internal prioritization processes (e.g., for FSAP)	Same as attributable performance indicator	Coverage of FSAPs compared to announced priorities	Internal reviews of FSAP initiative	Are effective priorities being implemented for undertaking FSAPs?	Low	Low	Medium

**THE QUESTIONS ASKED IN EACH OF THE TWELVE STEP OF THE REMLOGIC  
METHODOLOGY**

Step 1	How is IMF surveillance believed to work?
Step 2	Is the way it is believed surveillance works consistent with stakeholders' views, expert opinion and past evaluation findings?
Step 3	What routine monitoring can there be of whether the intermediate and final outcomes are being achieved (i.e. strategic indicators)?
Step 4	What routine monitoring can there be which attributes intermediate outcomes to IMF surveillance activity (i.e. performance indicators)?
Step 5	Assess how comprehensively the strategic and performance indicators cover the outcomes hierarchy
Step 6	What important evaluation questions can be asked for each intermediate outcome or group of intermediate outcomes?
Step 7	What evaluation questions are technically and practically feasible to answer?
Step 8	What is the likely cost of answering technically and practically feasible evaluation questions?
Step 9	What are priority evaluation questions within available evaluation resources?
Step 10	How can the evaluation questions be grouped into projects and phased?
Step 11	Collecting lessons from the priority evaluation projects
Step 12	Determining the next priority evaluation questions