



The Design and Implementation of a Menu of Evaluations



Gloria M. Rubio

Policy makers and program managers are faced every day with major decisions resulting from insufficient funding, ongoing complaints about service delivery, unmet needs among different population groups, and limited results on the ground. There is a menu of evaluation types implemented by developing and Organization for Economic Co-operation Development (OECD) countries to tackle a wide range of policy and program management issues, considering time, resources and capacity constraints. International experience highlights the importance of a gradual approach when introducing evaluation tools into country-level M&E systems. Different paths may work better for different countries depending on the main purpose of their M&E system, existing institutional capacity, the availability of funds, and external technical assistance.

Policy making and program management involve continuous decision making. Every day, policy makers and program managers tackle questions such as how to allocate resources across different interventions or program components, whether to add a new program or reform an existing one, how to choose between alternative service delivery mechanisms, or how to improve program operations. Any decision, even those favoring the status quo, entails some kind of assessment weighing different courses of action. To what extent these assessments are systematic and based on empirical evidence influences the chances of making a decision based on accurate information and being able to take advantage of opportunities to improve policy interventions.

Evaluation is the systematic collection and analysis of information to provide relevant feedback for decision making and policy formulation. There is a menu of evaluation types addressing a wide range of policy and program management questions. This note discusses the policy relevance, application, and requirements of different evalu-

ation types to facilitate the selection of a mix of evaluation tools depending on the information and feedback needs for decision making at any given moment in the lifespan of an intervention, while also accounting for time, resource, and capacity constraints. International experience illustrates uses of different evaluation tools across countries and over time.

Key Considerations in Selecting an Evaluation Tool

Evaluations are valuable when properly conducted and focused on their intended purpose. Hence, the selection of the right evaluation tool is critical. Choosing an evaluation tool, or combination of tools, depends on two main considerations. First, it depends on what information is needed to make major decisions. For example, is the evaluation prompted by the need to have overall performance information across programs to decide how to allocate resources? Or is it motivated by the need to have information on a particular program opera-

tion to identify bottlenecks that hinder adequate service delivery? Policy and program management issues and concerns drive the need for the evaluation, and in turn, point to the selection of a particular set of evaluation tools.

The second consideration in selecting an evaluation tool is the availability of resources to conduct the evaluation. The menu of evaluation tools includes different alternatives in terms of costs, duration, and capacity requirements. For example, some evaluations tools provide rapid feedback at relatively lower costs, while others require a longer implementation period, and their costs vary accordingly. Although there are usually trade-offs between evaluation depth and rigor and resource requirements, the selection of an evaluation type allows some flexibility in accommodating particular circumstances.

Key questions to ask before deciding on an evaluation tool, or combination of tools, include:

1. What is the purpose of the evaluation being conducted, that is, what decision will be informed by the evaluation?
2. What kinds of information are needed to make the decision? At what stage in the life of the program will the evaluation be conducted? These questions are closely related; information needs may vary depending on the program life stage. For example, a program that has just been planned would require an ex ante cost-benefit analysis to inform the decision on whether or not to implement it. Alternatively, a newly implemented program would benefit from information on how well operational procedures are followed, and whether there are any adjustments necessary for successful program operation.
3. How quickly or (unexpectedly) is the information needed? Sometimes evaluations are planned well in advance and results are expected in due course, giving maximum flexibility in evaluation tool choice. In other cases, the selection of an evaluation tool is influenced by the time frame of information demands from a particular policy process, such as budget preparation. In many cases, however, information needs arise suddenly, triggered by events that demand a quick response, such as a macroeconomic crisis or reform opportunity.

4. Who is the audience for the evaluation information? Some evaluations may be demanded internally within the organization by program management or policy makers in the ministry. Others may be externally required by the ministry of finance, congress, or a donor organization. Internal and external information needs may be different, leading to different evaluation types.
5. What resources (financial and human) are available to conduct the evaluation? Evaluation scope and tools should be aligned with available resources. In some cases, there may be an experienced team in charge of evaluations with an earmarked budget. However, in most cases, resources devoted to evaluation are limited or nonexistent, and institutional capacity is just developing.

Three Major Types of Program Evaluation

Once the key questions above have been answered, an evaluation tool or tools can be selected. Evaluations can be classified in several ways and some categories may overlap. For example, evaluations can be grouped according to the analytical paradigm (rationalist, pluralistic, participative, and so forth), purpose (formative versus summative), content (goals, process, outcomes/impact), time perspective (ex ante versus ex post), or the evaluator (internal versus external). Using a content-based classification, this section discusses three major types of program evaluation.¹ Table 1 summarizes the policy relevance, application, and requirements of each type of evaluation.

Goal- or Objective-Based Evaluations

Goal-based evaluations assess the clarity of a program's objectives and its progress in achieving these objectives. Questions asked by this type of evaluation include:

1. How were program objectives established? Was the process effective and are the resulting objectives clear and appropriate?
2. Is there a sound theory of change behind the program design?
3. Are program inputs, activities, and outputs aligned with the objectives or outcomes (internal consistency)?

Table 1. Main Evaluation Types

	Goal-based evaluations	Process-based evaluations	Impact evaluations
What are the main questions answered by this type of evaluation?	<ul style="list-style-type: none"> • Do policies or programs have clear objectives? • Is the program/policy design appropriate to achieve the objectives? • To what extent have policy/program objectives been achieved? 	<ul style="list-style-type: none"> • What are the actual steps and activities involved in delivering a product or service? How close are they to agreed operation? Are they efficient? • Are there adequate resources (money, equipment, facilities, training, and so forth) and systems (financial, management information, and so forth) in place? • Are program participants receiving quality products and services? 	<ul style="list-style-type: none"> • How have participants' well-being changed as a result of the program/policy? • Are there any unintended consequences, positive or negative, on program participants? • Are program/policy costs justified in terms of welfare impact?
When can this evaluation be conducted?	It may be conducted at early stages of program/policy implementation, for mid-term review, or at project completion	It may be conducted at any time, once or regularly, to check that implementation is on the right track, or when there are specific operational concerns	They are most effective when applied selectively to answer strategic policy questions or to assess innovative pilot interventions testing a new, unproven, but promising approach
What data collection and analysis are required?	Desk review of existing documents, possibly complemented by monitoring data analysis, interviews, and field visits	A mix of interviews with program staff and clients, user and facility surveys, record review, direct observation, focus groups, and analysis of monitoring data	Statistical and econometric analysis of survey and administrative data, ideally combined with qualitative data analysis
What skills are needed?	Policy and program analysis, possibly simple quantitative methods	Process analysis, quantitative and qualitative methods	Statistical and econometric analysis, possibly qualitative methods
How long does it take?	1– 3 months	1–6 months	<ul style="list-style-type: none"> • At least 6 months (retrospective evaluation) • At least 12–18 months (prospective evaluation)
What are the costs?	Generally low	Low to medium	Medium to high
What are some examples of this type of evaluation?	Mexico: Performance Specific Evaluation World Bank: Project Performance Assessment	<ul style="list-style-type: none"> • Process evaluations • Spot checks • Operational audits 	See suggested Web sites in Notes section
Additional comments		Large variation in depth and breadth	Time duration and costs depend on outcomes of interest and data availability

Source: Compiled by author.

4. Are there sound performance indicators to assess the program's progress in achieving its objectives?
5. What is the status of the program's progress toward achieving the objectives?
6. Will the goals be achieved according to the timelines specified in the program implementation or operations plan? If not, why?
7. Do personnel have adequate resources (money, equipment, facilities, training, and so forth) to achieve the goals?
8. Do priorities need to be changed to ensure focus on achieving the goals?
9. Should any goals be added or removed? Why?

This type of evaluation may be conducted at any time during program implementation. In early stages it could provide useful feedback on program design, such as early warnings on lack of clear objectives or inconsistencies between resources, activities, and objectives. It may also be a useful tool for a mid-term progress review or project completion assessment. This type of evaluation is usually based on a desk review of program or project documentation and may be complemented by analyzing monitoring data (if available). In some cases, this evaluation includes interviews with program staff to better understand the goal-setting process and the enabling or hindering factors behind observed progress. These are generally low-cost evaluations that can be completed in a short period of time.

Examples of this type of evaluation include:

- Mexico, Design Evaluations: This evaluation analyzes new programs' design, including their link to national development objectives, internal consistency, alignment between program log frame and operation rules, and potential synergy or duplicity with other programs. It relies on a desk review of program documentation, including the log frame.²
- Mexico, Performance-Specific Evaluation—This evaluation assess social programs' progress in meeting their objectives and annual targets based on a desk review of the Performance Evaluation System (SED) data. These evaluations are conducted annually for about 120 different programs and cost on average US\$5,500.³
- World Bank, Independent Evaluation Group, Project Performance Assessment—These assessments evaluate the relevance of projects' objectives, whether goals are being met, the ef-

iciency in their achievement, results sustainability, and institutional development impact. They are generally conducted at the end of the project by reviewing project documentation, visiting the borrowing country, and interviewing staff and government officials.⁴

Process-Based Evaluations

Process-based evaluations are aimed at understanding how a program works: What are the actual steps and activities involved in delivering a good or a service? How close are they to agreed operation? Is program operation efficient? There are numerous questions that might be asked in a process-based evaluation, including:

1. Is the program being implemented according to design?
2. Are operational procedures appropriate to ensure the timely delivery of quality products or services?
3. What is the level of compliance with the operations manual?
4. Are there adequate resources (money, equipment, facilities, training, and so forth) to ensure the timely delivery of quality products or services?
5. Are there adequate systems (human resources, financial, management information, and so forth) in place to support program operations?
6. Are program clients receiving quality products and services?
7. What is the general process that program clients go through with the product or program? Are program clients satisfied with the processes and services?
8. Are there any operational bottlenecks?
9. Is the program reaching the intended population? Are program reach-out activities adequate to ensure the desired level of target population participation?

This type of evaluation is useful in a variety of circumstances. For example, process-based evaluations are helpful to obtain early warnings of operational difficulties in newly implemented programs or components, particularly among those involving complex procedures. It may also be conducted at regular intervals to check that operation remains on track and follows established procedures, or at any time when there are staff or client complaints about service delivery. Furthermore, there may be fears of inefficiency

in long-existing programs that warrant a process-based evaluation.

Process-based evaluation tools vary considerably in depth and breadth, and thus in their costs and time required. They range from qualitative assessments in a few program sites to user or facilities surveys involving representative samples. Process-based evaluations use a combination of data collection and analysis methods including: interviews with program staff and clients; user and facility surveys; focus groups; direct observation; record review; and analysis of monitoring data.

Impact Evaluations

An impact evaluation assess if a program is producing the intended outcomes through providing relevant, quality outputs and services to the targeted population. Outcomes are the changes in well-being experienced by individuals as a result of participating in the program. They are different from program outputs, which are measured by the number of goods or services delivered or the number of people served. Outcomes involve beneficial transformations in participants' knowledge, attitudes, values, skills, behaviors, condition, or status. Intended outcomes vary depending on program objectives and design, for example, increased literacy, improved nutrition, decreased disease incidence, and so on. The main questions asked in an impact evaluation are:

1. Does the program or policy have the desired effects on individuals, households, or institutions?
2. Are these effects attributable to the program or would they have occurred anyway?
3. Are there any unintended consequences, positive or negative, on program participants?
4. Are program costs justified in terms of its welfare impact?
5. When there are various program implementation alternatives, which one is the most cost-effective?

Impact evaluations are useful to inform a range of policy decisions, from scaling up effective interventions, to adjusting program design, to curtailing unpromising interventions. In addition, they help generate evidence on which approach is more effective in reaching a particular objective when comparing different programs or different intervention options within a program.

Impact evaluations rely on quantitative methods (statistics and econometrics), but can

be complemented with qualitative analysis. Compared to other evaluation types, impact evaluations require more time, technical skills, and are costlier. Hence, they are most effective when applied selectively to answer strategic policy questions or to assess innovative pilot interventions testing a new, unproven, but promising approach. Although the impact evaluation process can start very early in a program life, results are not usually available until after several months of program implementation. Different outcomes of interest require varying exposure and maturation times for measurement. For example, changes in household consumption can be measured after six months of program implementation, but changes in nutrition outcomes require a longer period.

Impact evaluation results combined with program cost analysis create an additional evaluation tool. Cost-effectiveness analysis compares the relative performance of two or more programs or program alternatives in achieving a common outcome. Once impact and cost information are available across a variety of programs, cost-effectiveness analysis will allow policy makers to make informed decisions on which intervention to invest in.

Resources for impact evaluations include:

- The World Bank [Impact Evaluation Web site](#), which includes methodological and implementation guidelines for conducting impact evaluations as well as a database of evaluations of World Bank-supported interventions.
- The [Poverty Action Lab at MIT](#), which promotes the use of randomized evaluations to answer questions critical to poverty alleviation by providing methodological guidance, a randomized evaluations database, and policy lessons from cost-effectiveness analysis.

In addition to the three types of evaluation discussed here, there are other analytical tools and approaches that are widely used. Box 1 discusses briefly cost-benefit analysis and box 2 provides a glance to participatory evaluation.

Selecting a Mix of Evaluation Tools: International Experience

A number of monitoring and evaluation (M&E) systems at the country level use a combination of evaluation tools. The combinations vary across countries and have evolved over time. Country experiences show that the selection of an evalu-

Box 1. Another Analytical Tool

Cost-benefit analysis is a widely used tool to assess the desirability of a particular intervention. It answers questions such as whether a new airport should be built or a highway expanded. It compares the present value of the proposed intervention's benefits and costs. It is typically conducted *ex ante* and involves calculating intervention benefits in monetary terms and detailed cost estimates. It can be conducted at a high or low level of rigor, and costs vary accordingly. Examples of agencies using this technique include the Departments of Transportation in the United Kingdom, the United States, Canada, and the European Union.

Source: Author's compilation.

ation tool responds to differences in the driving principles of their M&E systems: transparency and accountability; policy and spending effectiveness and efficiency; or improved program management. The type of tool selected also reflects myriad country conditions, such as institutional capacity, resource availability, and institutional arrangements. Most countries have followed an incremental approach in the use of evaluation tools.

In Chile, for example, the Budget Department (Dirección de Presupuesto, DIPRES) at the Ministry of Finance is in charge of the Management Evaluation and Control System (MECS). Its main emphasis is to promote the efficient allocation of public spending, and it uses various M&E tools that were gradually added between 1994 and 2002. DIPRES made a conscious decision to start simple by incorporating the monitoring of performance indicators into the budget process. In 1997, DIPRES began conducting a relatively low-cost and rapid evaluation known as the Government Program Evaluation (box 3). Since then, these evaluations have been consistently used to inform the budget process. In 2001, impact evaluations were added to the Chilean M&E system. Programs are selected for an impact evaluation if they involve a substantial amount of public resources and there is no evidence from previous evaluations of their final outcome. DIPRES and the Congress jointly decide which programs should be evaluated each year.

In Mexico, the M&E system stems from an increasing demand for transparency and account-

Box 2. A Participatory Approach to Evaluation

Goal-, process-, and outcome-based evaluations can be conducted using a participatory approach.

This entails:

A process of self-assessment, collective knowledge production, and cooperative action in which the stakeholders in a development intervention participate substantively in the identification of the evaluation issues, the design of the evaluation, the collection and analysis of data, and the action taken as a result of the evaluation findings (Jackson and Kassam 1998).

Participatory evaluation emphasizes the voices and decisions of program participants when analyzing implementation difficulties or program effects, or when information is needed on stakeholders' knowledge of program goals or their views of progress.

Participatory evaluation usually employs rapid appraisal techniques, which are simpler, quicker, and less costly than other conventional data collection methods. This evaluation tool relies on qualitative and quantitative research techniques. Qualitative methods used include community evaluation committees, community workshops, self-directed focus groups, popular theater, community radio, transect walks, and wealth ranking, among others. Quantitative methods include short surveys. The choice of methods depends on local conditions, skills, and interests of various stakeholders.

Source: Jackson and Kassam (1998).

ability coupled with the demonstration effect of PROGRESA's (the Education, Health, and Nutrition Program) impact evaluation.⁵ Starting in 2000, the Appropriations Bill required an annual external evaluation for all federal programs subject to rules of operation. At the time, evaluation guidelines did not distinguish between different types of evaluations. However, every year external evaluations were expected to answer questions ranging from compliance with operation rules, to program impact, to program cost benefits. This was an overwhelming demand given the nascent evaluation institutional capacity, both from the demand (ministries) and the supply

Box 3. Overall Performance Evaluation

This tool is a combination of the three types of evaluations discussed earlier—goal-, process-, and outcome-based evaluations. A number of countries have incorporated this tool in their M&E system because it enables a rapid assessment of program performance through a desk review of program or project documentation, existing evaluations, and monitoring data. Using standardized evaluation criteria, overall performance evaluations can be compared across different programs and used for benchmarking.

This evaluation may be conducted at any time during the life of an intervention, although some assessment topics may have to be adapted for recently implemented programs. The main strengths of this evaluation are its short time requirement—three to five months—and relatively low cost. Its main disadvantages compared to more rigorous approaches are the weaker empirical evidence on which it is based and the typically very limited statistical or econometric techniques used for data analysis.

A number of countries include overall performance evaluations in their M&E systems:

- Chile, **Government Programs Evaluations**: Their purpose is to assess program design, operation, and results to inform the budget allocation process. They are based on a desk review of program documentation and staff interviews.
- United States, **Program Assessment Rating Tool**: Although it is no longer used in the United States, this tool served as a model for overall performance evaluations in countries such as Colombia and Mexico. This tool was used to evaluate a program's purpose, design, planning, management, results, and accountability to determine its overall effectiveness. Every year, 20 percent of agency programs were evaluated and all programs were evaluated at least once every five years.
- Colombia, **Executive Evaluations**: They provide feedback for mid-term program adjustment by analyzing programs' design, inputs, operation, results, organizational structure, strategic planning, and monitoring and evaluation.
- Mexico, **Consistency and Results Evaluation**: At the beginning of the new government administration, all federal government programs had an overall performance evaluation. This served as a baseline performance measurement and helped identify the strengths and weaknesses across programs and sectors.

Source: Compiled by author.

side (universities and research institutions). Then, in 2007, the newly created National Evaluation Council (CONEVAL), the Ministry of Finance (MoF), and the Ministry of Public Management (MoPM) jointly issued a revised set of guidelines including a set of evaluations tools and the criteria for tool application. In addition, the guidelines established the publication of an Annual Program Evaluation (PAE) that specifies which programs are required to have a particular type of evaluation, depending on their strategic relevance, particular policy interest, and previous evaluation results. The selection of programs and evaluation tools is made jointly by CONEVAL, MoF, and MoPM. Ministries can also propose additional evaluations to support policy or management decisions or to get specific program performance information.

Colombia's M&E system, SINERGIA, was launched in 1994. Its main focus is on national planning and accountability objectives. The system's original design included a monitoring scheme and a program evaluation component; however, the evaluation component did not become operational until 2002. Since then, SINERGIA's two main components consist of a system of performance indicators that tracks progress against the National Development Plan goals and an agenda of rigorous impact evaluations. Technical assistance and funding from donors allowed Colombia to opt for a more complex and costlier evaluation tool. Recently Colombia incorporated a lower cost and more rapid alternative known as Executive Evaluations (box 3).

Key Lessons

Using the appropriate tool, evaluations can help address a number of different management issues. The appropriate evaluation tool depends on the intended use of the evaluation findings, budget availability, time available, and capacity constraints. There are usually trade-offs between evaluation depth and rigor and resource requirements. Goal-based evaluations are quick and low cost, but tend to have data limitations and less rigorous analysis. By

contrast, impact evaluations are more rigorous, but require considerable technical expertise, take longer to yield results, and cost more. Moreover, it is important to realize that any particular type of evaluation can be conducted at a higher or lower analytical level, and thus costs vary accordingly.

Another important lesson is that evaluations are not a once in a lifetime exercise, nor does the use of one evaluation type exclude the rest. Policy and program information needs are continuous and change over time, so can the evaluation tool. Moreover, evaluation tools are complements rather than substitutes. When they are combined, they provide an even deeper understanding of programs and policies strengths and weaknesses. For example, whereas impact evaluations can produce reliable estimates of the causal effects of a program, they are not typically designed to lend insights into program implementation. Monitoring data and process evaluations are needed to track program implementation and examine questions of process that are critical to informing and interpreting the results from impact evaluations.

International experience highlights the importance of gradually incorporating evaluation tools into country-level M&E systems. Countries will require different evaluation tools depending on the main purpose of their M&E system, existing institutional capacity, and the availability of funds and external technical assistance. In some cases, using less complex evaluation tools, but ensuring they are properly applied and providing consistent feedback, may be an appropriate strategy to build an M&E system. In other cases, a few properly conducted and highly relevant impact evaluations with large demonstration effects may be a good way to motivate a shift toward results-based management.

Finally, as important as it is to select the right mix of evaluation tools, it is also important that evaluation results are used to inform policy and program decision making. Decision making informed and supported by reliable and systematic evaluations is more likely to lead to the success of policy and program interventions.

About the Author

Gloria M. Rubio is the former *Director General of the Social Programs Evaluation and Monitoring Unit at the Ministry of Social Development* in Mexico. Prior to joining the ministry, she worked

as *Research Associate at the Public Policy Institute in the Universidad de las Americas in Puebla*. She has collaborated with the *World Bank* as a consultant and staff member in Washington, DC. She holds a master's degree in Public Affairs from the *Woodrow Wilson School at Princeton University*. Currently, she works as a consultant, providing technical assistance in designing and implementing M&E systems and conducting analytical work on social policy issues.

Notes

1. There are other types of evaluations focused on other levels of public sector, including organizations and sectors that are not considered in this note.
2. See [Web site](http://www.coneval.gob.mx/coneval2/htmls/evaluacion_monitoreo/HomeEval-Monitoreo.jsp?categorias=EVAL_MON,EVAL_MON-diseno) for more information on Mexico's design evaluations, http://www.coneval.gob.mx/coneval2/htmls/evaluacion_monitoreo/HomeEval-Monitoreo.jsp?categorias=EVAL_MON,EVAL_MON-diseno.
3. See [Web site](http://www.coneval.gob.mx/) for more information on Mexico's performance-specific evaluation, <http://www.coneval.gob.mx/>.
4. See [Web site](http://web.worldbank.org/external/default/main?startPoint=0&startDate=2008%2F01%2F01&theSitePK=1324361&piPK=64254724&pagePK=64254514&menuPK=64253143) for more information on the World Bank's Project Performance Assessments, <http://web.worldbank.org/external/default/main?startPoint=0&startDate=2008%2F01%2F01&theSitePK=1324361&piPK=64254724&pagePK=64254514&menuPK=64253143>.
5. PROGRESA, now called Oportunidades, is a conditional cash transfer program launched in 1997 as an innovative pilot intervention to reduce poverty through human capital investment. The program impact evaluation became a role model for other social development interventions in Mexico.

Bibliography

- Departamento Nacional de Planeación. 2008. *Evaluación Ejecutiva Lineamientos Metodológicos*. República de Colombia
- División de Control de Gestión. 2009. *Metodología Evaluación de Impacto*. Gobierno de Chile, Ministerio de Hacienda, Dirección de Presupuestos
- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2010. *Impact Evaluation in Practice*. World Bank, Washington, DC.
- Jackson, Edward T., and Yusuf Kassam, eds. 1998. *Knowledge Shared: Participatory Evaluation in Development Cooperation*. Kumarian Press.
- McNamara, Carter. 2006. *Field Guide to Nonprofit Program Design, Marketing and Evaluation*, 4th Edition. Minneapolis, Minnesota: Authenticity Consulting, LLC.

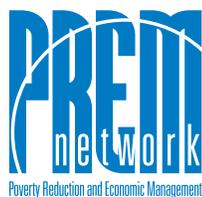
Zaltsman, Ariel. 2006. "Experience with Institutionalizing Monitoring and Evaluation in Five Latin American Countries: Argentina, Chile, Colombia, Costa Rica, and Uruguay." ECD Working Paper Series 16, Independent Evaluation Group, World Bank, Washington, DC.

Web Sites

Colombia, SINERGIA: <http://sinergia.dnp.gov.co/PortalDNP/>

Chile, DIPRES: <http://www.dipres.cl/572/propertyvalue-15697.html>

Mexico, CONEVAL: <http://www.coneval.gob.mx/>



This notes series is intended to summarize good practices and key policy findings on PREM-related topics. The views expressed in the notes are those of the authors and do not necessarily reflect those of the World Bank. PREMnotes are widely distributed to Bank staff and are also available on the PREM Web site (<http://www.worldbank.org/prem>). If you are interested in writing a PREMnote, email your idea to Madjiguene Seck at mseck@worldbank.org. For additional copies of this PREMnote please contact the PREM Advisory Service at x87736.

This series is for both external and internal dissemination