

Local Unit Alignment, Reorganization, and Consolidation Commission

Overview of the Literature Review and Analysis on Five Subjects Related to the Cost-Efficiency of Municipal Government

Marc Holzer, Ph.D.
Principal Investigator

John Fry
Project Director

Etienne Charbonneau
Deputy Project Director

Norma Riccucci, Ph.D.
Robert Shick, Ph.D.
Gregg Van Ryzin, Ph.D.
Supervising Faculty

Eileen Burnash
Librarian



May 6, 2009

TABLE OF CONTENTS

Overview of the Literature Review and Analysis on Five Subjects Related to the Cost-Efficiency of Municipal Government

EXECUTIVE SUMMARY	1
PURPOSE OF THE REPORT	3
CONCLUSIONS FROM THE REVIEWS OF THE LITERATURE	5
• From the Literature Review and Analysis Related to Optimal Municipal Size and Efficiency	5
• From the Literature Review and Analysis Related to Municipal Government Consolidation	7
• From the Literature Review and Analysis Related to Costs and Benefits of Service Delivery Consolidation among Municipalities	9
• From the Literature Review and Analysis Related to Optimal Service Delivery Arrangements and Efficiency	11
• From the Literature Review and Analysis Related to Measurement of Local Government Efficiency	13
AN INTEGRATED ANALYSIS OF THE CONCLUSIONS	15

EXECUTIVE SUMMARY

The School of Public Affairs and Administration (SPAA) at Rutgers University, Newark campus prepared five analyses for the Local Unit Alignment, Reorganization, and Consolidation Commission (LUARCC) and presented five reports early in 2009. This report analyzes the results of those literature reviews on municipal cost-efficiency in service delivery as affected by municipal size, municipal consolidation, service consolidation and alternative service delivery arrangements, as well as the review on the measurement of municipal efficiency. The major findings are these:

- Consolidation of municipalities has not been effective in increasing efficiency of municipal service delivery.
- The relationship between the size of the population served and efficiency is not consistent for the delivery of different service types.
- Size is associated with economies of scale for capital-intensive services and highly specialized services, resulting in greater efficiency in larger governments.
- Size is associated with diseconomies of scale for labor-intensive services, suggesting greater efficiencies in smaller governments.
- The most common measure of municipal efficiency is cost per capita.
- The literature considers the cost per capita metric greatly flawed due to inconsistencies of measurement and accounting in different jurisdictions.
- Performance measurement systems have addressed inconsistencies in measurement and accounting in order to provide valid comparative benchmarking.
- Benchmarking between jurisdictions has been effective in identifying performance improvements and cost savings.
- Data-based management review systems, such as the “CitiStat” or “CompStat” models, are effective at improving operational efficiencies.
- The perception is that sharing of services is effective in promoting efficiency of service delivery.
- Sharing and other forms of service consolidation are encouraged despite their logical reliance on economies of scale; the literature promotes sharing, even for services that demonstrate no economies of scale in studies of size and efficiency.
- Of the several modes of delivering municipal services, the most common is direct delivery by the municipality; the next most common is shared service delivery.
- Privatization and regionalization, provided by special districts or centralized governments, such as counties, provide increased efficiency in some implementations.

- Hybrid delivery arrangements, such as regionalized service delivery combined with a regional council or policy group, are appealing as alternatives to reduce the potential loss of local control that might result from consolidated service delivery arrangements.
- Hybrid forms of delivery foster the matching of different service delivery arrangements to the characteristics of specific service types, such as labor-intensive or capital-intensive services. For example, two-tier governments may provide regionally based water services through a special district, police patrol through the local municipalities, and economic development through the county.
- No service delivery arrangement is universally efficient, suggesting that a municipality always needs to complete a realistic, full cost evaluation of the projected results of an alternative service delivery arrangement before implementation.

Purpose of the Report

On March 15, 2007, New Jersey enacted the law creating a state commission, the “Local Unit Alignment Reorganization and Consolidation Commission” (P.L.2007, c.54). This law charged LUARCC with the responsibility to study and report on the structure and functions of county and municipal government. The study of the appropriate allocation of service delivery from the standpoint of efficiency was among the many responsibilities assigned to LUARCC.

An unprecedented special legislative session convened in June 2006, with the goal of enacting property tax reform, resulted in the enactment of this law. The legislature formed four bicameral, bipartisan joint committees, one of which was the Joint Committee on Government Consolidation and Shared Services, following Governor Corzine’s address to a joint session of the Legislature. The creation of a permanent commission (LUARCC) to facilitate municipal consolidation and the sharing of services among local units was one of the recommendations of that Joint Committee.

A principal assumption of the legislation creating LUARCC is that there are inefficiencies associated with the multiple municipal governmental units that exist in the State. The legislation suggests that municipal consolidation, particularly directed at the smaller municipalities, will result in more cost-effective government.

Accordingly, the legislature charged the commission with developing recommendations to improve the delivery of local government services. These recommendations may range from suggestions to share specific services to the realignment of local government units, including possible municipal consolidations, the merging of specific existing autonomous agencies into parent municipal or county governments, or the sharing of services between or among municipalities and other public entities.

As a step toward fulfilling this charge, the commission determined it would conduct a systematic review of the literature to determine the extent of existing data, studies, theories, gaps in the literature, and conclusions concerning whether consolidation of municipalities results in a more cost-efficient government. Following the issuance of scopes of work for five separate projects, the Local Unit Alignment, Reorganization, and Consolidation Commission (LUARCC) contracted with the School of Public Affairs and Administration (SPAA) at Rutgers University, Newark campus beginning October 6, 2008 to conduct a review of the literature on five different subjects related to the efficiency of local government.

The five reports submitted to LUARCC are:

- Literature Review and Analysis Related to Optimal Municipal Size and Efficiency
- Literature Review and Analysis Related to Municipal Government Consolidation

- Literature Review and Analysis Related to Costs and Benefits of Service Delivery Consolidation among Municipalities
- Literature Review and Analysis Related to Optimal Service Delivery Arrangements and Efficiency
- Literature Review and Analysis Related to Measurement of Local Government Efficiency

Conducting the Five Studies

Each of the five reports describes the methodology employed for the five literature searches in detail. The overall strategy was the same for each subject. Focusing on the academic literature, specific search terms for each of the subjects were applied against approximately thirty bibliographic databases and other information resources. Senior research associates, under the direction of faculty supervisors, evaluated the initial sources found in the search to determine which were appropriate to include in the study. The researchers examined and annotated the sources that were selected. Findings in these sources, which were appropriate to LUARCC's research questions, were recorded and used to create each of the five reports that resulted from these studies.

An important part of the work products specified in the contract is the annotated bibliography. Each of the five reports contains an annotated bibliography, with only the references appropriate to the subject of the individual report. The entire annotated bibliography for all five studies contains over 250 resources. SPAA has provided LUARCC with a copy of the software used to compile this bibliography (EndNote®, a high-quality bibliographic software program by Thompson Reuters) and the associated database created for this project, which includes the specific findings in each of the references.

The full text file from the original source is accessible through EndNote® or independently, usually as a PDF or Word document. In a few cases, a web link is provided or the reference is to a book, which is not included in the group of files provided to LUARCC.

Conclusions from the Reviews of the Literature

Each of the five reports commissioned by LUARCC presented the major conclusions from the literature, as expressed by the purpose of the literature review stated in the scope of work on that subject. This summary report lists the conclusions from each of the five reports containing the literature reviews.

From the Literature Review and Analysis Related to Optimal Municipal Size and Efficiency

Although the literature does not reveal strong and consistent relationships between size and efficiency in the delivery of local government services, some relationships are evident. The first relationship describes the overall effect of size of government on efficiency. The second and third statements show that the general relationship does not hold when specific services are considered. The fourth point, about the difficulties of measuring efficiency, is very important in the determination of what promotes municipal efficiency:

- **The Inverted U-shaped Curve**

There is an inverted U-shaped relationship between size and efficiency on a general level. Efficiency increases with population size up to about 25,000 people, at which point it is stable until size is about 250,000 people, and efficiency declines with increasing population size after that. The inverted U-shaped curve that describes the relationship between municipal size and efficiency offers two opportunities for improvement: the very smallest and the very largest governments. The literature defines the smallest as populations less than 20,000 to 25,000. The largest are municipalities with populations in excess of 250,000, which is a limited pool in New Jersey. However, in the one reference from the literature search that specifically studied New Jersey (Gabler, 1969), the diseconomy at the largest end of the spectrum was more uniformly pronounced than in two other states to which it was compared. The literature suggests New Jersey should examine its largest governments as well as its smallest governments for ways to increase efficiencies.

- **Service Specific Relationships**

The most important finding other than the inverted U-shaped curve was the difference in the relationship between size and efficiency in capital based services as opposed to labor-intensive services. Efficiency gains are related to size for capital or infrastructure intensive services such as sewer and water. The literature supports the finding that this same concept is operative for seldom used and specialized services, such as a high technology crime lab.

This suggests that contracting, sharing, or receiving specialized services from a larger entity can make selected services more efficient.

- Labor-Intensive Services

Labor-intensive services are more efficient in smaller governments. The literature only offers burdens of management control and excess administration in larger governments as an explanation for such inefficiencies. Reduced levels of services and expectations in smaller towns may also be operating to reduce costs. This finding of increased efficiency in smaller units is an important conclusion, because the literature attributes over 80 percent of governmental cost to labor-intensive services including police, fire, and education.

- Complexity of Measuring Efficiency

An additional finding from the literature is the difficulty in determining one measure of efficiency that works well at the level of a municipality or even for a service area. The most common basis for a measurement of efficiency is expenditure data, which is the numerator in the cost per capita indicator used throughout the literature. Varying definitions of expenditure measures across jurisdictions are the culprit in the comparability of this indicator. LUARCC has begun to evaluate the possibility of using municipal expenditure data in New Jersey to examine efficiency and the factors that promote it. While a difficult task, this makes a great deal of sense, because of the ability to attempt to control some of the confounding influences.

Finally, we must recognize that there are many inconsistencies in the literature. Even the most consistent findings of the inverted U-shaped curve and the relationship for capital-intensive versus labor-intensive services have variations supported by some authors. For example, we could not reconcile the debate over police responsiveness and inefficiency in small versus large units. Different authors observed greater managerial efficiencies in small units, but others saw relative efficiencies in large units. The literature does not provide a high level of confidence for further action on a systematic and broad basis.

From the Literature Review and Analysis Related to Municipal Government Consolidation

If the question is “Do we consolidate and where?”, the literature review shows there is not adequate information to say “yes” or “no,” and there are certainly not concrete criteria to follow through on a “yes” answer. The literature indicates that governments have achieved only a small portion of the attempted consolidations, and the results of those achieved are mixed.

- **Cost Savings Are Not Assured**

The literature does not consistently support the general belief that increasing the size of municipalities will lead to cost savings. The overall inverted U-Shaped curve relationship between population size and costs per capita masks the details of size relationships that vary with the nature of the service being delivered.

- **Implementation of Consolidation Is Costly and Time Consuming**

Transitional costs can be substantial if the consolidated government requires new facilities or infrastructure. It will normally take years to complete a consolidation. It is a complex task to achieve, as opposed to simply assuming a better organizational culture and better procedures in the merged government.

- **Savings Only Result from Reductions in Resources**

There is a tendency to create a wage and benefit structure that rises to the highest wage levels of the pre-merged governments. The merger must be evaluated as to how positions, equipment, or facilities can be reduced, if not immediately, then in the long run, so that these savings will balance the costs of the transition. The on-going costs resulting from higher uniform service levels and wage and benefit structures among merged municipalities are another major consideration.

- **Most Consolidation Attempts Fail**

Governments have not achieved the majority of recent voluntary attempts at consolidation. Of those that have been achieved, the results, measured in terms of costs savings or improved quality or responsiveness of service delivery, are mixed.

- **Politics Is an Obstacle**

The literature most often cites political considerations, running the gamut from multiple concerns of the public to concerns of political parties about

shifts in the balance of power, as the most significant obstacles to consolidation

Evaluations of potential mergers must be specific to the needs and situations of the communities involved. The mere attempt to consolidate often focuses an evaluation on ways other than consolidation to achieve increased efficiencies, even when the merger of governments does not result. This “unintended consequence” has led to better arrangements for service delivery in many situations.

From the Literature Review and Analysis Related to Costs and Benefits of Service Delivery Consolidation among Municipalities

The literature supports the assumption that inter-governmental cooperation can provide cost and quality efficiencies. In addition to discussing the benefits of shared services and other forms of cooperation, the literature provides insights into how to increase the amount of service delivery consolidation:

- **Quality Improvements**

Most shared services success stories assess cost savings, but some of these cases state that improvements in service delivery and quality should be the primary motivator in forming shared services. Sharing seldom reduces quality of the service, and often sharing improves quality. Quality issues are the sole motivation for some arrangements.

- **Providing High Cost Service Improvements**

There are functional service areas, such as information technology, in which local government has made only limited investments. Traditionally, this is because the government perceives the costs to be unpalatable to the citizenry. Service sharing offers the potential for quality improvement in areas such as this, without incurring the skepticism of the public about expensive investments.

- **Case-by-Case Resolution of Obstacles**

Although there are hurdles to implementation of shared services, they are situation-specific, and should be assessed and resolved in forming an agreement. There are a great many successful shared service arrangements. Even the concern with loss of local control or local delivery of the service is an infrequently mentioned obstacle in the literature.

- **Public Safety Services**

Reports from around the country indicate a reduction in public safety costs is possible. The transfer of police, fire, and dispatch services to other units of government are common potential efficiencies cited in the literature. Police and fire are difficult services to consolidate, because of the public's concerns with safety (in terms of response times) and organized labor's concerns with employment security.

- **State Encouragement**

Evidence shows that governments want to do more service sharing than they actually accomplish. States are trying to encourage collaboration through

grant and education programs. Some have tried to launch more of an outreach program, and others have tried to create databases of needs to meet through sharing.

- Need for Focus

Successful shared services require planning. Probably the greatest obstacle to doing what everybody seems to want – shared services that result in economies – is not spending the time and detailed effort to focus on it. The fact that most sharing arrangements come from informal networks and contacts is testimony to the fact governments do not spend enough time thinking about how to be efficient. They need to be encouraged to invest time and resources to do so.

- Incentives to Productive Organizations

Grants, such as New Jersey made available to develop county offices dedicated to promoting shared services, create incentives and the focus to evaluate more cost-effective strategies. These state grants to counties are a good step in the direction suggested by the Somerset County Business Partnership's proposal to fund effective sharing organizations in order to continue their successful work.

- Positive Effects of Fiscal Stress

Fiscal and budgetary pressures also encourage cooperation between units of government. Conversely, aid or grants, which maintain a service, are disincentives to finding alternative service delivery arrangements.

Finding the right level of government to deliver a service is an important concern of LUARCC. In addition to the consolidation of services discussed in this report, the report *Literature Review and Analysis Related to Optimal Service Delivery Arrangements and Efficiency* discusses some other mechanisms to move service delivery to the appropriate government, including centralized government, special districts, and regional delivery structures.

From the Literature Review and Analysis Related to Optimal Service Delivery Arrangements and Efficiency

This report reviewed varied approaches to service delivery arrangements. However, it is not obvious from what is found in the literature how to compare these approaches, much less propose an optimal array of service delivery arrangements. This area of alternate structures for service delivery is new and evolving, which makes the task of developing concrete conclusions more challenging.

The literature highlights the following arrangements as providing efficient and effective solutions for delivering public services:

- **Contracted Services**

Shared services can increase efficiency while retaining local control. The numbers show that a private contractor may be more efficient only if the municipality can monitor the contract adequately.

- **Special Districts Provide Specific Service Expertise**

Special districts (authorities and fire districts, for example) provide a focused area of expertise, but possibly at the cost of greater fragmentation, redundancy of general administrative functions, and loss of accountability. Regionalized special districts mitigate the increase in fragmentation and redundancy, but make accountability more distant.

- **Centralized Services Provide Scale Economies**

County or regional service provision is attractive for those services (infrastructure related or highly specialized) that show economies of scale, but there may be a need to provide representation of the constituent municipalities in order to assure local control.

- **Joint Boards for Specific Functions**

Joint boards are effective for unique services that are best provided over a broader area (for example, land use planning and economic development) and may provide local representation in determining how the service will be delivered.

- **Regional Policy Groups Increase Local Control for Varied Services**

Regional groups that are policy setting are very effective in providing local representation in an effort to collaborate and cooperate in providing services, but they depend on other structures to deliver the service directly. The Council of Governments in Monroe County, New York is an example.

- Hybrid Arrangements to Promote Efficiency and Provide Local Control

Arrangements that combine adequate local representation with an array of service delivery mechanisms applied appropriately to each specific type of service may optimize the benefits of different delivery arrangements. Two-tier governments are a hybrid form. They often combine a regional board, composed of representatives of constituent municipalities, with local delivery for services such as police patrols and regional or county delivery for services like economic development and land use planning.

The literature indicates there is potential in each of these approaches. It is important to fit the mechanism for delivery to the service type. In his review of fifty years of evidence on the relationship between the structure and performance of local governments, Bish (2001) suggests:

... given the diversity of communities and local services, no single organization can perform all the tasks demanded of local government. Metropolitan areas composed of a multiplicity of local governments and production arrangements are more responsive to residents' needs and generally provide local government services at less cost than monolithic amalgamations.

From the Literature Review and Analysis Related to Measurement of Local Government Efficiency

Performance measurement, and the general attempt to measure efficiency of local government, is an increasingly important concern. Both internationally and with increasing frequency in the United States, local governments are implementing performance measurement systems to improve efficiency and outcomes. Overall, the literature points to promising tools or approaches:

- **Performance Indicators (Measures)**

Performance measures should cover the range of the processes being monitored, from workload and inputs to outputs and outcomes, with a consideration for external variables, which may affect performance potential. Outcomes are most important because they measure those conditions that are concerns of the public and help managers, who accept their responsibilities for outcomes, to think outside the box. One also needs to focus on inputs, internal capacities, and outputs. The literature contains many good measures.

- **Data Envelopment Analysis**

Analyzing performance data has advanced in recent years. The most powerful tool appears to be Data Envelopment Analysis, which accounts for multiple inputs and outputs and distinguishes between technical, scale, and allocative efficiency. DEA requires sophisticated analytic expertise to use appropriately. However, good management can occur with simple, even basic, forms of data analysis, if it is informed by an understanding of the processes being monitored.

- **CitiStat**

CitiStat is an example of a successful data review and performance management system, but only one municipality in New Jersey has implemented it. Since it relies more heavily on management review than on sophisticated analysis or benchmarking, it is hard to replicate throughout an entire state. CitiStat does not preclude sophisticated analysis, but it does not rely on it to achieve its internal management goals.

- **Benchmarking (Inter-jurisdictional Comparisons)**

Local governments should make comparisons in order to maximize improvement in the delivery of services. Benchmarking against the performance of other governments is often a popular choice, but there are difficulties in creating comparable cost-based data. When data is available, stakeholders and managers will benchmark.

- Transparency of Performance Data

Managers and local officials debate whether they should publicize all performance data, but the literature considers the benefits of transparency and critical analysis as overshadowing the concerns of those who fear public reaction.

AN INTEGRATED ANALYSIS OF THE CONCLUSIONS

The conclusions of the literature do not support consolidation as an avenue for reducing property taxes. Economies of scale do exist, but according to the literature only for capital-intensive or specialized services that are relatively infrequent at the local level. While economies of scale appear to have potential in general for very small municipalities, this advantage disappears and reverses for labor-intensive services such as police patrols. That is, larger municipalities have reduced efficiencies for labor-intensive tasks.

Although the literature shows that consolidation, or the full merger of municipalities, is not a favored option, other mechanisms for consolidating services are in common use. Consolidation of services does not draw the negative reviews attributed to formal consolidations that result in the dissolution of a municipality. It is worth noting that attempts at formal consolidation, when they are not completed, often lead to other forms of consolidation of services that are more acceptable. Several of the studies of consolidation point to increased sharing of services between governments, transfer of service to a county or regional level, and, on some occasions, to the development of a regional group with constituent representation from the participating municipalities that encouraged cooperation and sharing of services among the members.

What accounts for the acceptance of consolidation of services but not the consolidation of municipalities? Transition costs are greater and the disruption more complex when consolidating whole municipalities. Clearly, the legal and political hurdles of formal consolidation are great. The literature shows that “politics” is a major obstacle. Hence, it advises consideration of costs in terms of what the merged governments realistically can achieve in the face of potential changes caused by harmonizing wages and benefits, increasing uniformity of services, and merging of different facilities and cultures.

An “obvious” conclusion from the consolidation literature is that savings only result if there is a reduction in resources required. Successful reforms must use fewer employees, less equipment, or less costly facilities than were used in total in the pre-reformed governments. Some of the literature demonstrates the increased efficiency in the resulting governmental unit based on clear prior understandings about how they would achieve desired results.

The literature describes alternative arrangements for delivering services, but the comparison of these alternatives is not comprehensive and consistent enough to choose one or another mode of service delivery. Further, there is no common identification of a set of general conditions and factors one should use to point to a better choice in a situation. There is a broad menu of choices but no concrete basis on which to make the decision. The literature frequently states that quality and cost of services must be evaluated in terms of the specifics of the situation.

Direct service delivery by a municipality is the most common form of service delivery. The next most common is shared services, which is the “apple pie” of alternative

arrangements for service delivery. The literature is not clear on why this is such a popular alternative to direct delivery, although it clearly is. In fact, the literature expresses concern that the impact of sharing services is not more carefully measured. There are estimates of the impact of sharing, non-quantified statements of the success of the cooperative arrangements, and even after-the-fact assessments of cost reduction or avoidance, but there is little that provides enough confidence to say how much cost saving can be achieved under any general set of conditions. Possible results must be evaluated for each situation.

Although examples may exist of sharing arrangements that are not beneficial, the literature indicates that, all else being equal, sharing is beneficial in service quality and cost. The literature also states that local governments want to cooperate to a much greater extent. States have developed grant programs and technical assistance mechanisms to encourage cooperation, municipalities have created regional groups to promote shared services, and fiscal stress has led to increased sharing. The common thread in the attempts to increase sharing is to focus attention on it. The literature reports that informal mechanisms are the most common sources of sharing. However, regional groups that are formed largely to find common solutions to local problems have reported considerable success in fostering cooperation.

These regional groups are not, however, that common in New Jersey. The promise of the New Jersey COUNT program, the effectiveness of private-public partnerships and informal county associations, and the effectiveness of a few counties in encouraging municipal cooperation and service consolidations is worthy of attention. The climate of fiscal stress also contributes to elected and appointed officials talking to each other about how to do more with less.

Evaluating Service Delivery Arrangements for Efficiency

Cost per capita is by far the most popular metric used. On the other hand, the literature is largely negative about the validity of cost per capita as a measurement of efficiency. The problems with this measure include the inconsistencies of cost allocations and differences in how individual jurisdictions define a service.

These inconsistencies will be present in any analysis of existing models of service delivery in different jurisdictions, and may distort the results. A careful analysis of costs can control some of these inconsistencies.

The review of performance measurement systems demonstrates that the comparability issue remains, but efforts to measure consistently across different jurisdictions can reduce many of the sources of inconsistency. If such a system is in place, comparisons can be made between different delivery arrangements. This would be in addition to the other benefits of a performance measurement system, including the capability to evaluate and monitor the quality of service delivery in an on-going fashion.

In general, the literature review indicates that decisions based on data are possible and more practical when there is a measurement system. Furthermore, that availability of information has led to continual improvements in performance and the development of a culture of rational decision-making.

The Building Blocks of Service Delivery Optimization

The literature about alternate arrangements for service delivery can be confusing because of definitions and semantics. Terms like regionalization and consolidation of services are very general. Using more specific terms is helpful in order to understand the options and how local governments can combine them.

“Scale” is a term that appears frequently in the literature. Changes in geographic scope determine scale. The common method of enlarging geographic scope (and, therefore, scale) is to increase the area in which a service is delivered by a single organization. Municipal consolidation, shared services, centralization, or the creation of a special district that serves multiple jurisdictions can achieve greater scale. The logic is that the needs for some resources, such as administrative personnel and equipment, do not grow as quickly as the scale of services being delivered, resulting in increased scale efficiency. The literature is of mixed mind about whether this logic comports with reality.

There is also functional scope, or scope of services. Functional scope refers to how many different services are offered by the same organization. A municipality usually delivers a full complement of services. By having a large scope of services, some general government functions become more efficient than in a single service organization. A single service provider duplicates administrative efforts and, therefore, administrative and overhead costs. A competing view is that this specialization of focus and expertise provides greater efficiency through singular attention to one service. The literature points out, however, that specialization often leads to higher wage costs to pay for the specific expertise in the service area.

The literature also defines service as both service provision and service production. Provision is the responsibility to budget for and make decisions about the amount of service to be delivered, or provided. Production is the actual act of bringing the service to the user or public. Production is a direct act or transaction, which is driven and defined by the budget and other policies of provision. A municipality retains responsibility for provision of a service when it contracts the service production to another government or the private sector. But when a service is centralized to the county, or a special district is formed, the municipality transfers the entire service responsibility.

When the municipality no longer provides for the service, it no longer has control over the service and it no longer needs to budget for such service provision. An implication of not budgeting for the costs is that the municipality no longer needs the source of revenue to support the service. The municipality can reduce or eliminate those fees or taxes (depending on how the service was funded), but those costs will be transferred with the responsibility to the special district or county. Some governments, through the

cooperative creation of regional policy groups or joint boards, counteract the concern about loss of control caused by relinquishing the responsibility for provision. Conversely, policy setting and budgeting may rest solely in the hands of the county or special district, with no input from constituent municipalities.

Advantages of Different Service Delivery Arrangements

Direct Service Delivery

Direct service delivery is the baseline method of service delivery. It is also the most common method. Its advantages include local control and local response, ability to set appropriate service levels on a uniform basis, the development of local expertise in the service, and the availability of resources (employees, equipment, and facilities) that the municipality can use to provide delivery of other service types during periods of weak demand in the original service type.

The disadvantages of direct service delivery revolve around cost and efficiency. The logic is that scale reduces unit costs; the reality, however, may be different. A smaller organization may attempt to keep costs down based on policy and operational decisions. The implications for quality may or may not be important.

For example, response and availability are essential for crime response and prevention, or any service with unscheduled demand, but not for trash collection and tax assessment, services that the municipality schedules. Resource levels in a small operation may need to be higher to meet peaks in demand (and to counteract unscheduled leave), leading to excess capacity that is not used efficiently in weaker demand cycles. Policy and budget decisions may sacrifice this excess capacity, and therefore response effectiveness, to attain overall efficiency.

On the other hand, service levels in a larger scale operation may be inappropriate for some parts of the area, especially if service levels are made uniform across the entire service area. A larger service area is more likely to lack homogeneity of service demand throughout the area, perhaps related to varying population densities, a rural-suburban-urban mix, or some other factors affecting service needs.

Small communities will use the same personnel resources to provide a range of services. Wearing many hats, the employee of a small jurisdiction providing a service may not be as expert as a specialized employee in a larger entity. Citizens may find their service needs met and may be happier with the transaction simply because a familiar face has provided the service. On the other hand, the larger organization, which contains more expertise, may be better at meeting the uncommon needs of the public and will be more likely to avoid serious errors based on lack of experience with unusual situations (e.g., tax lien calculations and redevelopment financing). The benefits of greater expertise also apply to internal administrative services available to municipal operating departments. Specialized expertise available in a larger government may support internal offices in

being more efficient when that expertise is needed, but the larger administrative department may result in slow, bureaucratic response to routine internal service needs.

Most, but not all, alternative forms of service delivery are intended to provide efficiencies of scale. Although the literature was mixed on this point, efficiencies of scale were seen to apply uniformly only for capital-intensive services, such as infrastructure development and maintenance, or highly specialized services, such as a police crime lab. Most authors believed that diseconomies of scale were in evidence for labor-intensive services such as police patrol. With that background, the next sections review delivery arrangements that take advantage of efficiencies of scale, if they do exist for the specific service type.

Shared Service Delivery

After direct delivery, the next most prevalent service delivery arrangement is shared services, and contracting in general. A shared service is a contract between two or more governments. It is usually contracted for one service at a time, but multiple service type arrangements can be made between the same partners. Typically, one municipality (or other government) acts as the lead agency, directly producing the service for all the partners and supplying the administration necessary to support the service production. Joint production of the service is uncommon, but mutual aid in fire and EMS response is an example where there is no lead agency. Joint production agreements are often more informal.

The provision of the service remains the responsibility of each municipality in which the shared service is delivered. The providing agency must budget for the service as well as make policy about the level of service delivery. Usually the providing agency compensates, through the budget allotment, the producing agency. Policy decisions about the service delivery and service levels made by the providing agency must be consistent with the agreement between the partners. Terminating the agreement is usually straightforward, if policy choices create an inconsistency or performance of the producer is considered inadequate.

The logic is that the service will be delivered over a larger service area, which is comprised of all the municipalities included in the agreement. If economies of scale exist for the service type, all partners should realize efficiency gains. If the scale is great enough, it may be practical to recruit specialized expertise and to acquire specialized equipment to provide the service. An example is the hiring of sophisticated computer experts and the purchase of high-end network servers and communications equipment to provide records storage and retrieval for a consortium of municipalities.

The partner responsible for the service loses control over the actual operation of service delivery, perhaps causing the citizens of this municipality to expect the new service deliverer will not address their needs. In a good shared service relationship, either staff of the municipality responsible for providing the delivery, or a citizen, will contact the delivering partner when there is a service delivery issue (for example, a missed trash

pick-up). Because all partners are governments, the relationship usually shows sensitivity to public demands.

Privatized Service Delivery

In theory, the only difference between a shared service and privatization is that in the latter case the contractor is a for-profit or non-profit organization, not a government. The evidence of the cost savings of privatization in the literature is so mixed that the determination of relative efficiency is not useful on a general level, but can be assessed when entering into a contract. This suggests there are at least three potential comparisons of costs for most services: direct delivery, shared services, and privatized delivery. In general, the advantages and disadvantages of privatization, compared to direct delivery of the service by the municipality, are the same as for a shared service.

However, some differences are more likely to occur with a non-government contractor. The most serious is that the contract is often harder to terminate, even in cases of poor performance. In a related manner, monitoring the contract is more expensive. The contractor may not share the sensitivity to public demands that a government does, but may have more of a business motive for contract performance.

The contractor may have more expertise, in both personnel and equipment, for the service type because it may be an area of business concentration. On the other hand, if the contractor is only providing the service type for the one municipality (more likely with a non-profit contractor), economies of scale will not exist.

Special District, Non-regionalized, Service Delivery

A special district, whether regional or based in only one municipality, is formed to provide one, or perhaps two related, service(s). On the surface, it is difficult to make a case for a special district serving a single municipality, unless there are other factors. At a regional level, the numbers might be more attractive.

A special district serving only the residents of the municipality, or only a part of the population, if there are multiple special districts providing the service type within the same municipality, provides no economies of scale. The logic of the advantage of a special district is the concentration of expertise on one service type. A special district duplicates much of the administrative resources and, therefore, administrative costs that are already part of the municipality, which the special district serves.

Compared to the direct provision of service by a department of the municipality, the sole difference in a non-regionalized special district is the formation of a new government, perhaps with authority for taxing, setting rates, and incurring debt. A special district requires the creation of an additional governing body, elected or appointed, for the delivery of the service type. At best, the special district operates at arm's length from the municipality, and usually the public it serves. The literature is clear that special districts are increasingly under scrutiny and being disbanded, although they are still being created

in some localities. The literature was silent on the benefits of a non-regional special district.

Compared to a shared service, it is very difficult to terminate this type of delivery arrangement. There must be a legal act to dissolve the special district entity. The assumption of debts and assets is a difficult issue. In addition, a cloud of mismanagement often accompanies the dissolution.

Regionalized Service Delivery Using a Special District

Like any regionalized service, there is a potential benefit from economies of scale, where they obtain, in a regionalized special district. Compared to a non-regional special district service, opting out of the regionalized service is easier for an individual municipality, because the dissolution of the special district is not necessary. However, the distribution of debts and assets remains an issue. Otherwise, the discussion above applies, but the difference of scale, when a special district serves several municipalities, is a very significant consideration affecting efficiency results.

Centralized Service Delivery

Regionalized service delivery can also be accomplished through centralization. Most centralization occurs at the county level of local government. In the literature review, we differentiated between a shared service and centralized service by determining whether the constituent municipalities had the option to maintain their own service production for the service type being centralized.

In a shared service, the municipality and the county enter into a contract, which either entity can terminate under the conditions of the agreement. The municipality maintains the responsibility for service provision, turning over just the responsibility for service production to the county. It continues to budget for the service and make policy decisions about service levels and delivery, although the county must agree with the amount of compensation it receives and the service it delivers.

In a centralized service, the municipality transfers both provision and production responsibilities to the county. The municipality removes the service from the municipal budget and has no formal control over service delivery. It becomes a county function.

The county service offers economies of scale, if economies exist for this type of service. Because of scale, there is potential for increased expertise. In a centralized service, uniformity of service delivery across the county will usually result. This can lead to service equity (a positive) or inappropriate service levels (a negative), depending on the homogeneity of service needs. Local control is absent, except informally.

Responsiveness and quality are dependent on the effectiveness of the operation the county establishes, as the literature shows. The use of local or central delivery of the service does not directly determine the responsiveness and quality of the service.

Hybrid Service Delivery

The arrangements covered above are reasonably comprehensive as a discussion of modes of service delivery. However, the field is evolving as governments seek better methods of service delivery, sometimes developing new forms of government to offset disadvantages arising from an existing service delivery arrangement. Regional policy groups and joint boards are two such forms. They may take other names and have variant implementations throughout the country and abroad, but their essence is a mechanism to re-introduce local control into a service delivery option in which the municipality has relinquished responsibility for both service provision and production.

In their pure form, regional policy groups and joint boards transfer service production responsibilities to other entities, most often counties or regionally operating contractors. The focus of their concern is both policy setting for provision and oversight for production. The basic distinction between a regional policy group and a joint board is the scope of services dealt with. The report on *Optimal Service Delivery Arrangements* describes the Council of Governments in Monroe County, New York and regional economic development associations as examples of the two forms, respectively. The regional policy group is usually formed when the county or a large city has taken over production for many service types. The joint board is usually formed to determine service policy for one service type. The lines of this distinction are not rigid, and, in fact, variations of the legal authorities vested in such a government and the organization of that government are often crafted to meet a specific perceived need.

Both of these forms attempt to get representation in decisions on provision from the constituent municipalities involved. Although they are governments with governing bodies, they usually do not directly produce services. The literature notes examples of regional policy groups that have been very effective in encouraging innovation and cooperation in restructuring how to deliver services when widespread consolidation of services was seen as necessary.

Hybrid forms of government also include two-tier government structures. Although this is nothing more than a combination of local and regional delivery of services, it appears that the recognition of the need for a two-tier government is accompanied by an increased focus on the best way to deliver each service. From the statements in the literature, this is exactly what will be most effective in maximizing efficiency of government service delivery. Some two-tier governments cited in the literature have:

- Moved services needing a regional perspective (economic development) to a regional level
- Maintained labor-intensive and citizen responsive services at a local level
- Provided a regional council to encourage collaboration and cooperation, identifying optimal methods of service delivery, and providing local oversight of regional services
- Dissolved special districts that were poorly managed, moving the functions to a central or local level, as deemed appropriate

- Consolidated infrastructure intensive services previously delivered by multiple entities into single regional district authorities
- Developed revenue sharing schemes to finance services of regional importance
- Developed central facilities (such as vehicle fueling), taking advantage of bulk purchasing to service individual local departments

Because the literature so often states that specific conditions foster specific service delivery responses, hybrid service delivery arrangements should be encouraged, including the customization of hybrid forms that are suggested by careful evaluation. Hybrids appear to be effective and usually result from an evaluation of specific issues and conditions. They are appealing because they utilize the advantages of an efficient delivery arrangement, while mitigating the disadvantages that are observed or anticipated in the non-hybrid service delivery arrangement.