Timeliness and Efficiency

Improving Access to Specialty Care

Mark F. Murray, M.D., M.P.A.

Waits and delays plague health care systems worldwide. Surveys have demonstrated that wait times for most specialists exceed those for primary care practices, and, dependent on the location and the specific specialty, average five weeks for nonurgent appointments.[1] These delays lead to widespread dissatisfaction and mistrust on the part of patients, demoralization of staff and providers; adverse clinical outcomes, increased cost because of rework and redundancy, the overuse of precious resources to triage patients, high fail-to-keep appointment rates, and suboptimum revenue.[2–15] Typically, patients are seen, screened, and referred from other more primary venues of care, such as primary care, emergency department (ED), or an urgent care center, which determine that more specialized care is needed for diagnosis or treatment, yet an extended wait is a major barrier to care.[16–18]

In addition, with increased delay into any component of the specialty care practice, systemwide workload increases. The delay increases phone calls to primary care and to specialty care, increases the requirement for use of resource as "triage," and increases cancellations and no-shows as well as unnecessary visits. Although the most perceptible delay in specialty care is the delay for an initial appointment, specialists inhabit systems of care that are fraught with other patient delay: at the ED, from the ED to the hospital bed or intensive care unit (ICU), from the ICU to hospital bed, as well as delays at testing, procedure, and surgical venues.[19–21]

Delays occur as a result of a mismatch of demand for service and supply of service or as a result of flow variation on the demand or supply sides. Specialty care providers live in a world of "supply competition," with numerous activities and duties competing for limited provider supply. This

Article-at-a-Glance

Background: Waits and delays plague health care systems worldwide, and wait times for most specialists exceed those for primary care practices. In office-based practices, the provider office presence is not diluted by competing indispensable activities, and the demand for service is most often for a single type, or stream, of office-based appointment demand. In the more complex specialty practices, however, the demand streams for office visits and other services compete for provider time and dilute the supply of office visits.

Seven Flow Strategies, with a Focus on the Initial Appointment: Seven strategies for reduction of delay can be applied, not only at all steps in patient flow and for all demand streams but also at all steps (for example, office visit, diagnostic procedure, surgery, follow-up) and within all specialty care types and ranges of practice. Each specialty care practice will need to discover how to use the basic principles and implement customized solutions within its own unique environment. Although it is ultimately critical to eliminate the delays in all streams of service, the focus is on the application of change strategies at the initial step between primary care and all specialty care practice types. The strategies are (1) balance supply and demand at each step in the chain, (2) work down the backlog, (3) reduce appointment types, (4) independent contingency planning for all variation, (5) reduce the demand for visits, (6) increase the supply, and (7) improve the efficiency of the office work flow.

Summary: Specialists support various, distinct demand streams that require demand/supply balance to achieve optimal system performance. If demand/supply balance exists within any stream, waits can be minimized, and the practice can choose time frames within which to balance workload.

Table 1. Glossary of Terms

System	A series of processes ordered in such a way as to achieve an aim.
Closed System	A system that has a closed, fixed enrollment, where there is a mandatory care relationship.
Open System	A system where patients, primarily due to their insurance or coverage plans, have a choice, and may or may not choose to seek a care relationship with these particular doctors.
Process	A series of tasks ordered in such a way as to achieve a specific aim; a set of processes makes up a system.
Smooth	No delays and maximum value.
Demand Streams	Distinct types of demand that require a distinctly different supply to resolve; the distinctly different supply could be a different room, a different provider, a provider doing a different kind of work, a different amount of time, a different venue, or different equipment.
External Demand	Demand generated outside the practice from referring doctors or referring venues, or directly from that population. This is demand that the practice does not directly control.
Internally Generated Demand	Return visits. Demand generated from inside the practice.
Packaging	Appropriate work-up.
Balance Box	Time frame within which the practice chooses to achieve balance of demand and supply.
Common-cause Variation	Random variation that occurs randomly in a system; cannot be controlled.
Special-cause Variation	Variation that is not inherent to the system; can be anticipated and controlled.
Backlog	Reservoir of waiting patients. Work in progress.
Provider Office Supply	The amounts of time providers spend in the office. The resource (supply) left over after all other indispensable duties is accounted for.

article explores this competition and the deeper system effects of delay to show how the initial delay "into the specialty practice"—that is, associated with the patient's access to specialty care—can be reduced.

Delays in Specialty Care

Multiple interventions, primarily focusing on "scheduling systems," have been attempted, resulting in the proliferation of multiple appointment types and guidelines, rules for demand management, increased triage and gatekeeping, and the use of mid-level providers as intermediaries. None of these interventions have succeeded in significantly altering the patient experience of delay.[22–38]

Recently, Schall et al., using principles previously developed to address delays in primary care practices, demonstrated improvements in waits and delays in both primary and specialty care settings in Veterans' Administration (VA) practices.[16] Because of the specific nature of the VA environment—primarily salaried physicians and a "closed" system (Table 1, above), this work has not been considered universally applicable.

By using the same principles with modification for each specific environment, "open" systems—generally, health care organizations with nonsalaried clinicians and a wider range of patient choice about specialists—have attained even more impressive achievements in their efforts to reduce delays. These organizations, by focusing on the various duties competing for provider (supply) time, have developed a deeper understanding that the wait time between primary care and specialty care represents one wait in a series of system steps. Isolated optimization of this step could have an adverse effect on overall system performance by "solving" the wait at the initial step while pushing the wait time deeper into the system.

Access to care in the specialty arena requires looking at patients' initial passage into specialty care, as well as analyzing their entire journey across specialty care services. Although patients, referring providers, and specialty care staff are frequently frustrated by long waits at the initial care step, specialists themselves are often inundated with another set of competing tasks: procedures, testing, surgery—both operating room (OR) or ambulatory surgery center, and ED and "on-call" coverage. Performance of these duties often takes precedence over office-based appointments. Because delays for much of this work is deemed intolerable, specialists focus their attention on these deeper system responsibilities, relegating office visits to a lower priority. Thus, office visit appointments are the point in the system

Table 2. Why Do Queues Form?

Demand > Supply	When demand for service is greater than supply of service, a waiting time will ensue. Example: if daily demand is 10 units of service and daily supply is 10 units of service, there will be no waiting time. If daily demand is 11 and daily supply is 10, each day one more patient will wait.
Variation in Either Demand or Supply	If the average demand is 10 and the average supply is 10, but supply or demand is variable, a waiting time will ensue. If daily demand ranges between 5 and 15 and supply is fixed at 10, 5 units of supply go unused on days when demand is 5. On days when demand is 15 and supply is fixed at 10, 5 demand units wait. Because unused supply cannot be passed forward, variation creates an inevitable waiting time. Supply variation (e.g., the specialist cancels appointments due to emergency surgery) is more common than demand variation and creates most specialty care office delays.
Paradigm	In health care, an accepted paradigm goes: if the patient is really sick and can prove it, that patient is prioritized into a line with a short wait. Patients who cannot prove they are really sick are placed in a line with a longer wait. This prioritization inevitably lengthens delays for less acute patients.
Use of a buffer	In many specialty care environments, queues are used as a buffer for assurance of revenue or to guard against the risk of unused supply. While the attempt to get "100% utilization" seems efficient, this buffering strategy is often costly and wasteful, creating rework and redundancy, adding to fail to keep rates, and diverting resources to "triage" patients into urgent and nonurgent queues.

where the queues inevitably form. Reasons for the formation of queues are shown in Table 2 (above).

Demand Streams in Primary and Specialty Care

The single term *specialty care* obscures the wide range of dynamics at work in these practices. In office-based practices, the provider office presence is not diluted by competing indispensable activities, and the demand for service is most often for a single type of office-based appointment demand (single demand stream). Within that demand stream there is competition, of course, between new patient visits and return visits. In the more complex specialty practices, however, the demand streams for office visits and other services compete for provider time and dilute the supply of office visits.

The fundamental dynamic in primary care and the entire range of specialty care practices is the same: demand for services, whether these services are multiple or single, has to be balanced by a corresponding supply or resource for optimal system performance. The range of complexity is determined by the number of distinct types of work or demand streams addressed by the practice. Thus, some specialty care practices with a minimal number of distinct types of work (for example, dermatology with a primary focus on office workload and office procedure) act more like primary care than the more complex specialty care practices with more distinct streams of demand (for example, obstetrics/gynecology with streams of office work, "on-call" function, deliveries, OR, ambulatory surgery center, hospital, and other procedures.)

The provider office supply in specialty care, then, is the total provider supply minus the supply or resource required to support the nonoffice activities deeper inside the flow system. The competition between the various demand streams and even within the demand streams (new versus return, for example) compounded by demand variation, contributes to turbulence, unpredictability of flow, and delays. For long-term viability, any system requires a balance of the sum of the demand stream work with a sum of the supply available to support that demand. All the competing streams need an aggregate balance. Without that overall balance, and even if one stream is permanently imbalanced, an increasing wait time will be inevitable. At the same time, most specialty care practices do have an overall balance, as manifested by relatively stable but persistent wait times that fluctuate between the various and competing demand streams. For example, the delay for an appointment may be shortened but, at the same time, the delay for a surgical procedure will be lengthened, and then this situation reverses itself. The overall delay, however, remains lengthy, but persistent and stable, implying that overall demand is balanced by supply but the delivery of the service is delayed.

Practice complexity is directly related to the number and variety of competing demand streams. The strategies

for improvement in each stream are the same, although the specific applications of those principles often differ.

Seven Flow Strategies, with a Focus on the Initial Appointment

The seven strategies for reduction of delay, developed and derived from observation of such leading businesses as Toyota, Starbucks, and Amazon.com, and that match demand for service to supply of service, can be applied not only at all steps in patient flow and for all demand streams but also at all steps (for example, office visit, diagnostic procedure, surgery, follow-up) and within all specialty care types and ranges of practice. Each specialty care practice will need to discover how to use the basic principles and implement customized solutions within its own unique environment. Although it is ultimately critical to eliminate the delays in all streams of service, this article focuses on how the change strategies are applied at the initial step between primary care and all specialty care practice types. The specialty care delay problem cannot be definitively resolved by looking at only one step at a time, but the initial step from primary care to specialty care is often a good starting point because (a) delay is most easily viewed and measured at the initial step; (b) the greatest number of patients are delayed at that step; and (c) delays amplify as the work flows further inside the system, making it crucial to reduce the delay at the initial step.[39–42] For specialty care groups with provider supply devoted primarily to outpatient work, for example, dermatology and rheumatology, improving the flow of work at this step should solve the entire flow problem.

It is possible to solve the initial delay by temporarily pulling resources from other duties. For example, waits can be reduced for new and return orthopedic patients by adding more office hours and subtracting OR hours; the result is fewer delays in the office because of the favorable demand-supply balance but greater delay at the surgical step. Thus, improving the wait time at any step requires an understanding of the entire system flow dynamic. To improve the flow and reduce the waits across the system, each link in the chain needs to be evaluated, measured, optimized, and kept in constant balance. To achieve optimal system performance, it is critical to determine the linkages between the steps, understand and discover the system bottleneck, link all the steps, and smooth the entire flow. This process of delay improvement is thus iterative: back and forth between the initial step and the other steps. The seven change strategies that follow can be applied at each step in the patient's journey to improve flow.

1. Balance Supply and Demand at Each Step in the Chain

The first step in access improvement is to understand, measure, and achieve a balance between demand and supply at each step. Demand is commonly measured by looking at past activity, which represents a retrospective view of how much work was completed within a time frame. This may be different than the amount of work generated within the same time frame. Thus, demand must be measured prospectively, as it is generated.[2,3]

Demand at the initial step into specialty care has two components: externally generated demand (workload generated from outside the practice, either from patients or referring entities) and internally generated demand (work generated from within the practice as requests for return visits). Each of these demand streams must be measured, evaluated, and compared to the available supply of appointments.

To understand and influence external demand, it is important to stratify this demand into the following components:

a. *What is the work?* There are various categories of work sent to specialty providers, based on diagnosis, symptom, or condition.

b. *Who sends the work?* There will be variation between one primary care provider and another in terms of the amount of type of work sent.

c. *Where does the work come from?* The packaging of the work often differs significantly depending on the venue (for example, from the primary care provider versus the ED).

For specialty care practices with stable and consistent daily provider presence, it is possible to match the appointment demand with the appropriate appointment supply each day. This single-day balance box is achievable if providers are present > 60% of the time and are reasonably distributed over that time frame. For most specialty practices, however, the office supply is diluted to ≤ 50% of the total potential office time and is spread unevenly across the week. In addition, other factors are present: (a) the geographical distance of the referring venue from the specialty care practice; and (b) the need for more information, tests,

or procedures before the visit, which create delay for the initial specialty care visit. These factors create tension for an extended time frame; thus, most practices will widen the supply-demand balance box to five days. The system dynamics—steps required to achieve balance and deal with variation—remain the same, but instead of the primary care mantra of doing "today's work today," the goal is to do "this week's work this week." If there is a balance between demand and supply, this stability can be achieved within any chosen time frame. If there is an imbalance in any of the demand streams, however, waits will inevitably get worse, and nothing can solve the problem.

The distinct types of demand that come into specialty care practices include demand for office appointments (which can be subdivided into new patients and return patients), surgery in the OR, surgery in an ambulatory surgery center, procedure or test, hospital consultation, and the catch-all miscellaneous demand captured by the "on-call" function. These demand streams are managed by an allocation of resource or supply. Most practices use intuition, experience, and some data, as well as a concern about clinical issues to ensure that the supply is sufficient to meet these various different types of demand. The priority in the competition between demand streams is based on acuity. A demand/supply balance has to be achieved within each demand stream and across demand streams. Total demand (all demand types, converted to time and added) must equal all supply time. If one demand stream is mismatched, either permanently or temporarily, a delay will ensue. Eventually the wait gets intolerable. Practices can, however, tolerate variation as long as in the long run the demand equals the supply. This temporary mismatch is managed by tolerating a wait and then catching up. The more acute demand streams (on-call function and hospital) are always balanced without fail, and the least acute (office, primarily the return appointment demand stream) are the most tolerable and absorb the variation. For example, if half the specialty providers are not working in any venue within the practice, some demand streams are prioritized and supported at 100% (the on-call function and hospital) while the office, with low priority, is supported at far less than 50%. Hence, the longest and most variable waits are seen in the office. If this wait is temporary because of a temporary demand/supply mismatch, a practice can "get away with it," but if it is caused by a permanent mismatch, then the wait will just get worse, leading to practice system breakdown. Even though this article focuses on the office work alone, it should be noted that it is critical to measure the demand and supply in all the competing demand streams. In the office, if the total appointment demand generated during a five-day period is equal to appointment supply for the same period of time but there is an appointment delay, then the next strategies employed include the strategy to work down the backlog of work and strategies to reduce the variation in demand or supply to achieve a smoother flow and less fluctuation with these temporary delays. Palo Alto Urology and Camino Medical Group Otolaryngology (Sidebar 1, page 130, and Sidebar 2, page 131) are particularly strong examples of measurement and review of the allocation of supply for the correct ratio of supply to balance the measured demand for that service.

2. Work Down the Backlog

Backlog reduction is a necessary strategy to recalibrate a delayed system of care. In specialty care, backlogs of patients often exist at every system step. Wherever a demand/supply mismatch is present, either permanent or temporary, a delay ensues. These delays need to be measured, evaluated for stability, and then stabilized. If wait times are worsening, demand reduction and supply enhancement strategies (as described below) need to be employed first to achieve stability. Once stabilized, the backlog can be eliminated and the system recalibrated.

To eliminate backlog, a period of time must be spent completing more work than is generated. The requisite extra supply can be added within the daily schedule or as a bolus of supply somewhere within the week. Some specialty practices face a larger backlog for new patients, whereas others have a larger and expanding backlog for return patients. Both streams need to be measured and evaluated for delay and demand/supply balance. To achieve optimal system function, the new patient demand stream, in which delay and dissatisfaction occur most frequently, needs to be balanced first.

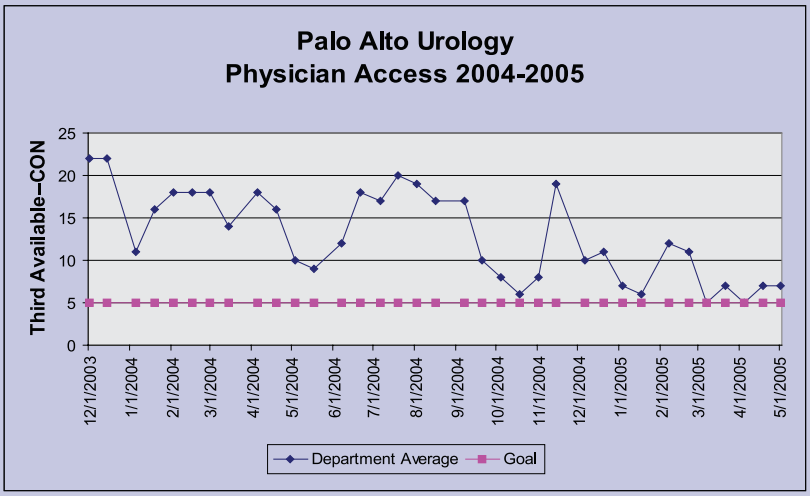
In addition to adding supply, the following enhancement strategies will indirectly help reduce backlog:

a. Getting the referred patient on the right route for care and ensuring that the appropriate specialist is consulted to reduce secondary referrals for care and ensuring that patients see the same provider of care for subsequent visits[43,44]

Sidebar 1. Palo Alto Urology

Palo Alto Medical Group, a division of Palo Alto Medical Foundation, is a multispecialty medical practice staffed with 410 physicians within 34 departments. The urology group has been working on access improvement for 2 years.

Synopsis of results: Achieved aim of reducing wait time to 5 days or less. Improved referral provider satisfaction. CON, consult.



Key Changes (selected as specific changes from the high-leverage change ideas):

- Measured wait time, appointment demand, and relative supply
- Reviewed how specialists allocate time
- Reduced backlog
- Measured office lead time
- Synchronized appointment times
- Developed service agreements with primary care
- Improved "graduation" of patients back to primary care

b. Maximizing the efficiency of each visit and doing more within each visit

c. Altering return visit rates[45,46]

d. Developing appropriate care delivery models, wherein patients not appropriate for referral back to primary care can be seen by nonphysician resources within the boundaries of the specialty care practice[47–51]

e. Streamlining and improving the flow of work across the office so that, with the same intensity of effort, more patients can be seen and evaluated

f. Reviewing all venues of work (procedure, hospital, surgery) for inefficiencies—if the flow at other venues can be improved, supply can be gained to add support to office practice.

Backlog reduction requires the support of leadership, which must aid in the development of measurement guidelines, set dates for the start and completion of backlog reduction, add support staff to aid in the work, and align the incentives so that providers who reduce their backlog early do not receive overflow from other providers. The Camino Medical Group Otolaryngology (Sidebar 2) practice had a strong backlog reduction plan.

3. Reduce Appointment Types

Reducing the appointment types in specialty care to a minimum number is one of the most powerful delay reduction strategies. An appointment type is a specific visit type that has both inclusion and exclusion criteria. A "new patient" appointment is often distinctly different than a "return follow-up patient" appointment. Pushing any work to the future, through use of multiple appointment types and their resulting queues, creates inflexibility in the future schedule, rigidity inside the scheduling system, and the necessity for triage, which uses up resources. It also increases rework, redundancy, and the likelihood of no-shows, and inevitably leads to longer waiting times.[2–3,18] Reducing appointment types allows any patient to be seen in any appointment slot, thereby reducing variation and delay.[52]

Although the goal is to minimize appointment types to achieve less patient waits, some distinct appointment types are necessary to maintain smooth flow. Any specialty care practice with a balance box larger than 24 hours, a low new-to-return patient ratio, or provider office presence of < 50% will need distinct new and return appointment slots. For example, in a strictly obstetrics practice, the ratio of new patients to returning patients is low, and the providers are often in the office < 50% of the office time.

In this setting, a distinction between new and return appointments is necessary because without that distinction, the return patients would crowd out the new patients. In addition, distinct appointment types must be created when there is a clear need for a specific specialist, a specific room, specific support staff, a specific time frame, or specific equipment.

Although the new and return appointment types will compete within the office demand stream, from the patient and referring provider customer viewpoint, the most critical wait time is that for new patients. In specialty care practices where there is > 50% provider absence from the office because of support of numerous other demand streams, then "pooling of new patient referrals," that is, sending the work nonspecifically to the department, allows appointments to be made with the first-available new-patient appointment slot. If work is sent to specific specialty care providers in a predetermined way (nonpooled), then if a provider is absent from the office for extended periods, longer waits will inevitably result.

Sidebar 2. Camino Medical Group Otolaryngology

Camino Medical Group is a multi-specialty medical practice located in San Jose, California, staffed with 223 physicians within 46 departments. The otolaryngology group has been working on access improvement for 18 months.

Synopsis of results: Achieved aim of reducing wait time to 5 days or less. Improved referral provider satisfaction.

Key Changes (selected as specific changes from the high-leverage change ideas):

- Explicit backlog reduction plan
- Service agreements with primary care
- Standardized rooms
- Use of physician's assistant for follow-up and minor procedures
- Prepare in advance for visits

4. Implement Contingency Planning for All Variation

If a specialty care practice is committed to reducing the waits to within a specific time frame, it will need to plan for all demand and supply variation. In most practices, demand for new appointments is relatively stable and merely needs to be measured. Internally generated demand (return work) shows more variation because of specific provider practice styles.[45] Contingency planning requires an understanding of this variation, and, while averages of demand and supply are useful metrics, fine-tuning is necessary. Statistical process control graphs can be used to view the variation in the demand for service and the supply of resource and can help identify causes of this variation. Special cause variation (artificially created variation) must be addressed and eliminated, and common cause variation (normal expected variation) must be understood and stabilized. Improvement strategies include measurement and planning for any seasonal demand variation, bringing return discretionary visits back in times where there is less need for new visit time as a buffer to smooth demand, development of time-off policies to smooth the supply resource, and planning for providers returning from practice absences.

Equitable distribution of new patient workload among providers is another way to smooth demand. Making each provider responsible for a panel or caseload of patients helps create this equity. Panels or caseloads need not be equal but must be large enough to keep the provider busy and small enough that the provider can complete the work each day or within each week. To keep the wait time stable, providers must absorb demand variation on a daily or weekly basis; some days or weeks may be heavier than others.[2–3,16–17,42]

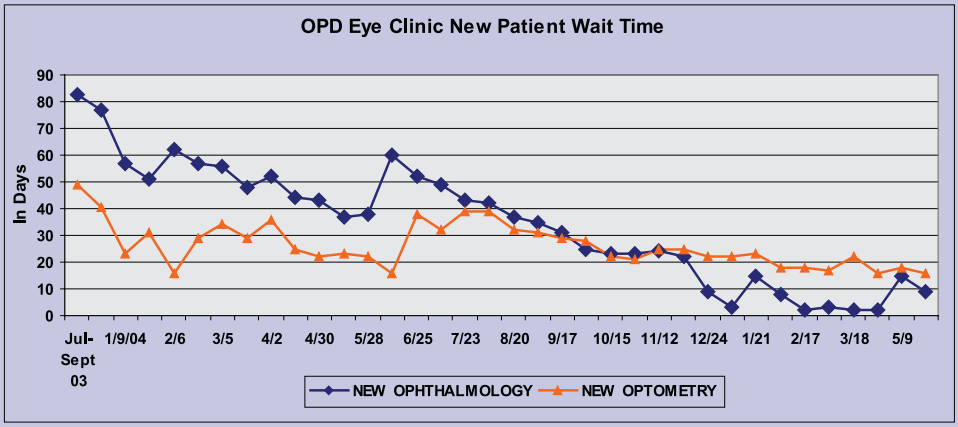
Specialists create unique value by seeing new patients; dividing new patients equitably among providers creates the tension to graduate patients (that is, refer them back to primary care) to open up space for the other new patients.[16–18]

Office waiting times in specialty care practices with multiple demand streams competing for allocation of supply often fluctuate widely because of prioritization based on acuity of the allocated supply. Because the work done in many practices (general surgery, for example) is done linearly and in sequence, that is, patients are initially seen, then evaluated and then sent deeper into the system for another service; if there is overallocation in one step and

Sidebar 3. Santa Clara Eye Care

Santa Clara Valley Medical Center, a public hospital owned and operated by Santa Clara County, is staffed by a 370-physician multispecialty medical group and supports an academic teaching program. The Eye Care service department recently completed its second year of implementation.

Synopsis of results: Achieved aim of reducing wait time to 5 days or less. Improved referral provider satisfaction. Reduced staffing cost.



Key Changes (selected as specific changes from the high-leverage change ideas):

- Measured wait time, appointment demand, and relative supply
- Reduced backlog
- Measured office lead time
- Synchronized appointment times
- Increased number of clinics by identifying additional space
- Expanded diabetes mellitus screening clinic because patients with diabetes were seen less efficiently in the general eye clinic
- Opened one clinic just for new patients

underallocation in another, temporary demand/supply mismatches result. This, in turn, creates boluses of work and work waiting. For example, if a surgeon planning an absence spends a disproportionate amount of time in the OR to reduce surgical backlog before that absence, backlog will still increase into the office—not only during but before the absence. On the surgeon's return, the first priority is often to "make up" on-call time, which in turn generates more OR cases and keeps the surgeon away from the office. The office backlog continues to build. Eventually, there are not enough surgery cases to fill the "block time," so the surgeon then withdraws from the OR and moves back to the office, where "hidden" surgical demand is uncovered within the backlog into the office. This in turns creates an extending wait for the OR. This wide fluctuation in under- or overallocation creates temporary mismatches in each of the various demand streams, and the back-and-forth movement creates boluses, batches, delays, system turbulence, and variation—all of which can be addressed by planning for these contingencies. For example, if the office provider supply is low, the wait time goal for new patients can be maintained by a higher ratio of new patients on each office provider's schedule. Alternatively, if office provider supply is high, then the new-patient ratio per provider can be lowered, allowing more return visits on the schedule. Measurement and a conscious plan to avoid wide swings in allocation are key.

5. Reduce the Demand for Visits

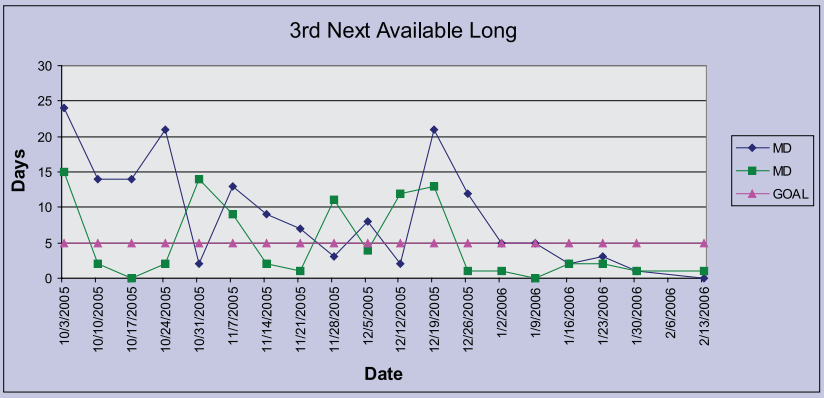
Because most work comes to specialists through a filter of primary care, the most effective demand reduction strategy is the development of service agreements.[16–19] A service agreement is an agreement between any two entities in a flow system, one of which sends work to the other.[19] It defines the proper work and outlines the correct packaging of that work, which directs the right work to the right person and ensures that it is prepared in such a way that it can be efficiently dealt with once it reaches the specialist. Service agreements also afford the opportunity to streamline the referral process itself, reducing or eliminating steps to reduce or eliminate patient delay. Service agreements between primary care and urologists, for

example, can define what work in the urology arena is done by primary care providers (monitoring of stable prostatic-specific antigen levels), what work is done by urologists (scrotal mass), and what is the proper packaging prior to referral (a set of defined tests for patients with microscopic hematuria). When providers recognize the commitment to stabilize patient wait times within a specific time frame, they then recognize the tension in the new-to-return-patient ratio. Because there is a limit to the potential provider supply, providers gain the incentive to optimize that ratio by graduating patients back to primary care or managing patients within the specialty practice with less direct physician contact. Other demand reduction strategies focus on continuity, individual provider return visit rates, and the use of technology.[44–46] These strategies work best in environments based on continuity, relationship, and trust, in which the patient realizes that they can see their own doctor at any time, for any problem. The Palo Alto Urology practice (Sidebar 1) developed a strong service agreement.

Sidebar 4. Marshfield Clinic Pain Management

Marshfield Clinic is the largest private group medical practice in Wisconsin, with more than 725 physicians representing 86 different medical specialties and serving over 360,000 unique patients. During a three-year period, Marshfield Clinic has been involved in an organization-wide effort to improve access to care for all patients to all specialties. The pain management practice has been involved in access improvement for 18 months.

Synopsis of results: Achieved access aim of offering 100% of patients an appointment within 5 days. Achieved office efficiency aim of 35% reduction of follow-up appointment cycle time. Eliminated costly overtime. Improved staff satisfaction with job.



Key Changes (selected as specific changes from the high-leverage change ideas):

- Measured wait time, appointment demand, and relative supply
- Commitment to complete all work that can arrive prior to 3:30 P.M. each day
- Reduced backlog by adding more appointments temporarily
- Standardized appointment slots
- Review future schedule on a daily basis
- Measured office lead time
- Synchronized appointment times
- Improved continuity
- Revised referral process
- Post-vacation contingency plans with "held" slots
- Revised provider schedule by adding slots when necessary and eliminating hidden time
- Standardized examination rooms and process

6. Increase the Supply

Practices can increase supply by adding more hours or more providers or by subtracting unnecessary work from the provider. Any flow process has a rate-limiting step, which in health care is the provider. The entire journey can move only as fast as the slowest step, that is, the greatest mismatch of demand and supply.[52] When patients traverse the specialty care network, the slowest step is commonly found in the office setting. Thus, increasing supply means driving all unnecessary appointment work away from providers for the provider to be freed to perform work for which she or he is uniquely qualified. Hematology-oncology practices have used this strategy for years by having much of the follow-up chemotherapy work done by specialized practice nurses and not physicians. This in turn requires a care team task and workflow analysis to ensure that the right person is doing the right work.[47–51] By using a retinal camera, the Santa Clara Eye practice

(Sidebar 3, page 132) was able to leverage the physicians' time and functionally increase the supply devoted to direct patient care.

7. Improve the Efficiency of the Office Work Flow

Improving the workflow within the office setting itself requires an explicit plan to address patient delays during the appointment. The plan includes flow-mapping the patient's journey, identifying the value-added steps, and eliminating non-value-added steps and waits between steps. Streamlining the patient's journey by reducing delays for the patient-provider interaction frees more time for that valuable interaction. With the same effort, specialty providers can see more patients within the same time frame. This does not mean working faster; it means that distractions and interruptions are eliminated. Streamlining the flow of work also requires synchronizing the work: consistently getting the patient, provider, information, equipment and staff to an open room on time.[53,54] The Marshfield Clinic Pain Management practice (Sidebar 4, page 133) focused attention on synchronizing the daily appointment workflow and preparing in advance for that work.

Summary

Specialists support various, distinct demand streams, all of which require a demand/supply balance to achieve optimal system performance. Each stream is linked operationally and can be linked with measurement. Movement of resource from one stream to another disrupts the delicate balance. Thus, measurement and balance of all streams is necessary. If there is a demand/supply balance within any stream, waits can be minimized, and the practice can choose the time frame within which to balance the workload. Although a complex choreography is at play, the seven flow strategies will ensure that waits and delays are reduced or eliminated. J

The author would like to acknowledge the work, dedication, and effort of the following individuals in improving access to specialty care at their own sites and in contributing to the article: Marshfield Clinic Pain Management, Edna Devries, M.D., Linda Pelton, Heidi Reigel; Camino Medical Group: Phil Brosterhouse, M.D., Nina Quintal; Palo Alto Medical Group, Jenny Buchanan, Susan Smith, M.D.; Santa Clara Eye Care, Kent Imai, M.D., Chris Snow, M.D., Donna Decarlo, Joe Eliason, M.D.

Mark F. Murray, M.D., M.P.A., is Principal, Mark Murray & Associates, Sacramento, California. Please address correspondence to murraytant@msn.com.

References

1. Merritt, Hawkins & Associates: *2004 Survey of Physician Appointment Wait Times.* Irving, Texas: Merritt, Hawkins & Associates, Summary Report: 1-12, 2004.
2. Murray M., Berwick D.M.: Advanced access: Reducing waiting and delays in primary care. *JAMA* 289:1035–1040, Feb. 26, 2003.
3. Murray M., et al.: Improving timely access to primary care: case studies in the advanced access model. *JAMA* 289:1042–1046, Feb. 26, 2003.
4. Kennedy J.G., Hsu J.T.: Implementation of an open access scheduling system in a residency training program. *Fam Med* 35:666–670, Oct. 2003.
5. Boelke C., Boushon B., Isensee S.: Achieving open access: The road to improved service and satisfaction. *MGM Journal* 47:58–68, Sep.–Oct. 2000.
6. Valenti W.M., Bookhardt-Murray J.: Advanced-access scheduling boosts quality, productivity and revenue. *Drug Benefit Trends.* 16:510,513–514, May 2004.
7. Lacy N.L., et al.: Why we don't come: Patient perceptions on no-shows. *Ann Fam Med* 2:541–545, Nov.–Dec. 2004.
8. Moore C.G., Wilson-Witherspoon P., Probst J.C.: Time and money: Effects of no-shows at a family practice residency clinic. *Fam Med* 33:522–527, Jul.–Aug. 2001.
9. Barron W.M.: Failed appointments: Who misses them, why they are missed, and what can be done. *Prim Care* 7:563–574, Dec. 1980.
10. Hixon A.L., Chapman R.W., Nuovo J.: Failure to keep clinic appointments: implications for residency education and productivity. *Fam Med* 31:627–630, Oct. 1999.
11. Belardi F.G., Weir S., Craig F.W.: A controlled trial of an advanced access appointment system in a residency family medicine center. *Fam Med* 36:341–345, May 2004.
12. O'Hare C.D., Corlett J.: The outcomes of open-access scheduling. *Fam Pract Manag* 11:35–38, Feb. 2004.
13. Carlson B.: Same-day appointments promise increased productivity. *Managed Care* 11:43–44, Dec. 2002.
14. Giannone J.: Open access as an alternative to patient combat. *Fam Pract Manag* 10:65, Jan. 2003.
15. Herriot S.: Reducing delays and waiting times with open-access scheduling. *Fam Pract Manag* 6:38–43, Apr. 1999.
16. Schall M., et al.: Improving patient access to the Veterans Health Administration's primary care and specialty clinics. *Joint Comm J Qual Saf* 30:415–423, 2004.
17. Duffy T.E.: Urology advanced clinic access concepts. Paper presented at the 4th Annual International Summit on Redesigning the Clinical Office Practice, St. Louis, Apr. 14, 2003.
18. Parenti C., Pierpont G., Murray M.: Reducing wait times for cardiac consultation. *Federal Practitioner* 22: pp. 24–28, 31, Feb. 2005.
19. Murray M.: Reducing waits and delays in the referral process. *Fam Pract Manag* 9:39–42, Mar. 2002.

References, *continued*

20. Asplin B.R., et al.: A conceptual model of emergency department crowding. *Ann Emerg Med* 42:173–180, Aug. 2003.
21. Forster A.J., et al.: The effect of hospital occupancy on emergency department length of stay and patient disposition. *Acad Emerg Med* 10:127–133, Feb. 2003.
22. Berry L.L., Seiders K., Wilder S.S.: Innovations in access to care: A patient-centered approach. *Ann Intern Med* 139:568–574, Oct. 7, 2003.
23. Bodenheimer T.: Innovations in primary care in the United States. *BMJ* 326:796–798, Apr. 12, 2003.
24. Dewitt P.: Finding the time: Clinics pursue new strategies to reduce wait time for appointments. *HealthLeaders* 11:72–73, Sep. 2004.
25. Kilo C.M., et al.: Improving access to clinical offices. *J Med Pract Manag* 16:126–132, Nov.–Dec. 2000.
26. Kofoed L., Ramirez M.E.: Improving access: A model for mental health care. *Federal Practitioner* 21:11–12,17-20,23,26, Sep. 2004.
27. Murray M.: Modernizing the NHS. Patient Care: Access. *BMJ* 320:1594–1596, Jun. 10, 2000.
28. Murray M., Tantau C.: Must patients wait? *Jt Comm J Qual Improv* 24:423–425, Aug. 1998.
29. Murray M., Tantau C.: Same-day appointments: Exploding the access paradigm. *Fam Pract Manag* 7:45–50, Sep. 2000. http://www.aafp.org/fpm/20000900/45same.html (last accessed Jan. 4, 2007).
30. Berwick D.: As good as it should get: Making health care better in the new millennium. Paper presented at the National Coalition on HealthCare, Institute for Health Care Improvement, Boston, Sep. 1998.
31. Kilo C., Endsley S.: As good as it could get: remaking the medical practice. *Fam Pract Manag* 7:48–58, 2000.
32. Singer I.: Advanced access: A new paradigm in the delivery of ambulatory care services. Paper presented at the National Association of Public Hospitals and Health Systems, Washington, D.C., Oct. 1, 2001.
33. White B.: Starting a revolution in office-based care. *Fam Pract Manag* 8:29–35, Oct. 2001.
34. Randolph G., et al.: Behind schedule: Improving access to care for children one practice at a time. *Pediatrics* 113(3 pt 1):e230–e237, Mar. 2004.
35. Murray M., Tantau C.: Redefining open access to primary care. *Managed Care Quarterly* 7:45–55, Summer 1999.
36. Smith J.: Redesigning health care. *BMJ* 322:1257–1258, May 26, 2001.
37. Murray M.: Waiting for Healthcare: Physician offices can dramatically reduce how long patients wait for appointments [editorial]. *Postgrad Med* 113:13–14, Feb. 2003.
38. Kolata G.: Harried doctors try to ease big delays and rushed visits. *The New York Times:* Jan. 4, 2001. http://www.nytimes.com/2001/01/04/science/04DOCS.html (last accessed Jan. 4, 2001).
39. Cawley P., Hanlon P.: Hospital medicine programs add value to the throughput process. *The Hospitalist* 8:11–14, Sep. 2004.
40. Rozich J.D., Resar R.K.: Using a unit assessment tool to optimize patient flow and staffing in a community hospital. *Jt Comm J Qual Improv* 28:31-41, Jan. 2002.
41. Shea S.S., Senteno J.: Emergency department patient throughput: A continuous quality improvement approach to length of stay. *J Emerg Nurs* 20:355–360, Oct. 1994.
42. Murray M.: Answers to your questions about same-day scheduling. *Fam Pract Manag* 12:59–64, Mar. 2005.
43. Plauth A., Pearson S.D.: Discontinuity of care: Urgent care utilization within a health maintenance organization. *Am J Manag Care* 4:1531–1537, Nov. 1998.
44. Raddish M., Horn S.D., Sharkey P.D.: Continuity of care: Is it cost effective? *Am J Manag Care* 5:727–734, Jun. 1999.
45. Petitti D.B., Grumbach K.: Variation in physicians: recommendations about revisit interval for three common conditions. *J Fam Pract* 37:235–240, Sep. 1993.
46. Schwartz L., et al.: Setting the revisit interval in Primary Care. *J Gen Intern Med* 14:230–235, 1999.
47. Grandinetti D.A.: Make the most of your staff. *Med Econ* 8:56, Apr. 2000.
48. Grumbach K., Bodenheimer T.: Can health care teams improve primary care practice? *JAMA* 291:1246–1251, Mar. 10, 2004.
49. Patel V., et al.: The collaborative health care team: the role of individual and group expertise. *Teach Learn Med* 12:117–132, Summer 2000.
50. Weymier R.E.: Ideas for optimizing your nursing staff. *Fam Pract Manag* 10:51–52, Feb. 2003.
51. Wagner E.H.: The role of patient care teams in chronic disease management. *BMJ* 320:569–572, Feb. 26, 2000.
52. Hall R.: *Queuing Methods for Services and Manufacturing.* Englewood Cliffs, N.J.: Prentice Hall, 1991.
53. Carlson B.: Working too hard, doctor? Poor work flow could be to blame. *Manag Care* 11:48–49, Jul. 2002.
54. Endsley S., Magill M.K., Godfrey M.M.: Creating a lean practice. *Fam Pract Manag* 13:34–38, Apr. 2006.