Improving Clinical Access and Continuity through Physician Panel Redesign

Hari Balasubramanian, PhD¹, Ritesh Banerjee, PhD², Brian Denton, PhD³, James Naessens, ScD², and James Stahl, MD⁴

¹Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst, MA, USA; ²Division of Health Care Policy and Research, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA; ³Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC, USA; ⁴Department of Medicine, MGH Institute for Technology Assessment, Massachusetts General Hospital, Boston, MA, USA; ⁵Present address: Analysis Group, Inc., Boston, MA, USA.

BACKGROUND: Population growth, an aging population and the increasing prevalence of chronic disease are projected to increase demand for primary care services in the United States.

OBJECTIVE: Using systems engineering methods, to re-design physician patient panels targeting optimal access and continuity of care.

DESIGN: We use computer simulation methods to design physician panels and model a practice’s appointment system and capacity to provide clinical service. Baseline data were derived from a primary care group practice of 39 physicians with over 20,000 patients at the Mayo Clinic in Rochester, MN, for the years 2004–2006. Panel design specifically took into account panel size and case mix (based on age and gender).

MEASURES: The primary outcome measures were patient waiting time and patient/clinician continuity. Continuity is defined as the inverse of the proportion of times patients are redirected to see a provider other than their primary care physician (PCP).

RESULTS: The optimized panel design decreases waiting time by 44% and increases continuity by 40% over baseline. The new panel design provides shorter waiting time and higher continuity over a wide range of practice panel sizes.

CONCLUSIONS: Redesigning primary care physician panels can improve access to and continuity of care for patients.

KEY WORDS: primary care access; continuity of care; systems engineering.

J Gen Intern Med
DOI: 10.1007/s11606-010-1417-7
© Society of General Internal Medicine 2010

INTRODUCTION

The US faces a shortage of primary care physicians (PCPs), a result of increasing clinical demand in an aging population, and a shrinking number of PCPs likely to be in practice in the near future. Timely access and continuity of care, two key goals of primary care practices, have suffered. Insufficient primary care access will have serious consequences. For example, it is estimated that 40% of emergency department visits result from patients not being able to access their PCP in a timely fashion. From 1997 to 2001, the percentage of people reporting being unable to obtain a timely appointment with their PCP rose from 23% to 33%.

Continuity of care also suffers with fewer PCPs, and this too has serious consequences. Patients who regularly see their own PCPs are more satisfied with their care, more likely to take medications correctly and less likely to be hospitalized. Lack of continuity also reduces the effectiveness of care and can increase the number of follow-up appointments.

To address some of these issues, many primary care practices have tried implementing advanced access. Advanced access promotes the concept that physicians should “do today’s work today” rather than push appointments into the future. Because of the intrinsic variability of patient demand, the supply of physician time must be sufficiently greater than demand for advanced access to work.

Proponents of advanced access (AA) claim that since patients are offered appointments on the same or next day, wait times are shorter. Moreover, since AA works with an individual physician’s calendar, the patient tends to see their own PCP more frequently, facilitating continuity. However, several studies have documented significant barriers to successful implementation of AA. These difficulties include the difficulty working through existing backlogs; adequate follow-up care in panels with large proportions of chronically ill patients; and maintaining continuity of care, since prioritizing speed of access frequently comes at the cost of less continuity. Others have discussed problems in the interpretation of advanced access and suggested more research needs to be done to establish its effectiveness.

Whether a group practice adopts advanced access or not, any initiative for improving timeliness and continuity has to consider the size and composition of physician panels in the practice. How might the design of a physician’s panel affect the appointment burden of a practice? Size matters, but size alone is not the only factor. The number of patients in a
Panel and their disease burden (case mix) together determine the panel’s aggregate demand for access. Our hypothesis is that redesigning physician panels to account for case mix (age and gender) can reduce total patient waiting time and increase the frequency with which patients see their own provider. We test our hypothesis using a computer-based simulation model and, further, assess the effect of panel allocation by case mix on the effective capacity of the practice.

**METHODS**

**Baseline Data**

We collected appointment and physician availability data from the panels of a primary care group practice at the Mayo Clinic in Rochester, Minnesota, from 2004 to 2006. The practice consisted of 39 physicians and covers approximately 20,000 patients living in Olmsted and the surrounding counties. For our analysis, we grouped patients by gender and age, as suggested by Murray. Age was further subdivided into 14 age categories of 5-year increments starting at age 18 years old through age 83 years old, for a total of 28 patient categories. While more elaborate classification systems are available, we chose age and gender because these are the simplest indicators of appointment frequency that illustrate the benefits of our approach. Figure 1 illustrates the distribution of the fraction (or percentage) of total patients requesting appointments in a week for two categories—males (48–53 years old) and females (73–78 years old). The two distributions show how appointment request rates can vary with gender and age. The distributions are different with regard to both mean and variance.

The 39 physicians in the Mayo PCIM (Primary Care Internal Medicine) practice cover 20,000 patients. The practice group is equivalent to 17 physicians working full time, after accounting for part-time and other activities (e.g., education and research). The average panel size for this practice is approximately 1,200 patients per physician. To obtain panel sizes more representative of the typical practice (~2,000/provider), we inflated the total empanelled population to 34,000 by sampling, while keeping the proportion of people in the different demographic categories (i.e., case mix) the same. The FTE adjusted panel size thus increased from 1,200 per physician to about 2,000 per physician on average. The composition of patients in the new panels is unchanged relative to the original panels.

The practice does not use advanced access, but ensures that patients needing same day appointments (roughly 40% of total requests) are offered an appointment slot, not necessarily with their PCP. There is no limit on how far in advance patients book their appointments.

**Panel Redesign**

Panel design is an allocation problem: given a set of health categories (e.g., age, gender, comorbidities) and a given number of physician panels in a group practice, how many patients from each category should be allocated to each panel? Figures 2 and 3 show a conceptual illustration of panel redesign using three patient categories (based on age) and three physicians. We stress that this is merely an illustrative example; our actual model consists of 28 patient categories and 39 physicians as described above. The arrows indicate current panel allocations; the width of the arrows indicates the appointment demand. Physician C is overburdened because of her case mix; her requests are well over her available capacity. This will result in increased wait time and losses in continuity. In Figure 3, after reassigning patients from physicians C and B to A, the requests for all physicians are in balance with available capacity.

Redesigning physician panels can thus reduce workload while simultaneously improving access (equivalently minimize waiting time) and continuity of care. Variability also plays an important role. A physician overburdened with high-demand, high-variability patients (for example 73–78 F patients in Fig. 1) will experience unanticipated spikes in demand with the consequence that her patients will likely fail to secure a timely appointment and will tend to see other physicians. This result is in agreement with queueing theory, which tells us that waiting increases as the variability of demand increases. Panel redesign essentially reduces the negative effects of demand variability by redistributing part of the high-demand and high-
variability patients to physicians whose capacity profiles allow them to accommodate additional patients.

**Optimal Design**

To find the optimal design, we use stochastic linear programming\(^2\), which searches through the large number of panel allocation possibilities to find the best design. Such techniques are an important methodological area within the field of systems engineering. They are analytical tools that can be used to study planned changes rigorously before implementation and have been applied to problems in other service industries including the design of transportation systems and airline management.

---

**Figure 2.** Panel redesign example—part I. Conceptual example showing how panel allocations can result in mismatches between requested appointments and available capacity. Panel allocations are indicated with arrows; the width of the arrows indicates the appointment demand. Physician C is overburdened because of her case mix with the result that her requests are well over her available capacity. Physician A, on the other hand, has spare capacity.

**Figure 3.** Panel redesign example—part II. Conceptual example showing how panel allocations from Figure 2 are redesigned so that requests for each physician are in balance with available capacity. Some of Physician C’s patients are redistributed to the other two physicians (especially Physician A), with the result that her requests match with available capacity.
Baseline and Capacity-Based Design

We compare the results from our optimal panel design arrived at with two other panel strategies. The first, the baseline design, is the design currently used by the Mayo PCIM practice. The second is a capacity-based panel design.

The capacity-based design is a straightforward allocation strategy that allows us to evaluate the performance of a practice in which patient panels are balanced on average physician capacity. The capacity-based panel design is constructed as follows: first, we tabulate each physician’s average share of the total average weekly capacity (available appointment slots) of the group practice. For example, if physician “A” sees patients on average for 40 h a week out of a total of 200 h of patient-time by the group (five PCP practice), her share is 20%. Each PCP receives a proportion of patients from each patient category equal to the proportion of time he or she is available to see patients.

Simulation and Statistical Methods

Each of the three designs—baseline, capacity based and optimal—is evaluated over one year using a computer-based simulation, which uses the principles of queuing theory to calculate summary statistics such as average wait time and the number of redirections to other providers. The simulation mimics the practice’s appointment scheduling system. In each week, patients make appointment requests that are satisfied on a first-come-first-served basis. When a physician’s calendar in a week is full, patients can either choose to wait for a future week to see their own provider, or they can see another physician in the same week (provided capacity is available). If capacity is not available, extra slots are added to accommodate these patients. These extra slots represent the additional hours put in by a non-PCP physician to cover the demand.

In our model, our baseline assumption, based on the rate observed at the Mayo Clinic, is that 40% of patients when given the option will choose to see an alternative PCP now rather than wait. In essence, these patients may represent acute-care patients with immediate needs, for whom timeliness is more important than continuity. The remaining—presumably chronic care patients for whom continuity is more valuable—are willing to wait to see their own PCP. Those patients redirected to other physicians are subsequently redirected back to their own PCP rather than follow-up with the new PCP. This reflects the fact that seeing another physician generates additional follow-up appointments.23

To generate weekly demand in our simulation, we sample randomly (with replacement) from historical visit data (2004–2006 or 156 weeks) for each of the 28 age and gender categories (Fig. 1 presents examples of such data in histogram form). Each physician has a weekly schedule that we use to determine weekly capacity. The results we present are averages of 200 replications of the simulation for each design. Ninety-five percent confidence intervals of the measures (wait time and number of redirections to non-PCP physicians) are constructed for each design based on the replications.

In addition to wait time and continuity, we also report the total utilization of the clinic, the number of total slots filled over total slots available. The number of extra slots (additional capacity) that the clinic needed on a weekly basis to meet demand are included in this calculation. For both these measures, we provide averages based on the 200 replications and 95% confidence intervals.

Adding New Patients to the Practice

As the demand for primary care doctors increases in the US, practices are routinely faced with decisions regarding whether to empanel new patients. Young patients (less than 35 years old) of either gender tend to use appointments less frequently than older patients. For our first sensitivity analysis, we increased the proportion of young patients by 25%, keeping everything else the same. This increased the total patient pool by 3,500 patients. We then analyzed how the different panel designs perform under this scenario.

Adding New Patient Categories

Though age and gender are good proxy measures for case mix, more specification can be useful. We used Classification and Regression Tree (CART) Analysis24 to identify conditions in addition to age and gender that are significant predictors of visit rates. Our analysis of the clinic data revealed that coronary artery disease (CAD), hypertension and depression were strongly predictive of visit rate. In conjunction with age and gender, we identified 15 categories based on these factors that could be used to categorize patients. We tested panel
redesign under this new classification to determine the sensitivity of our model to the type of patient classification.

RESULTS

Baseline, Capacity-based and Optimal Design

For the base-case scenario the mean waiting time was 0.57 weeks (4 days), and there were on average 266 redirections to other physicians per week. For the capacity-based design scenario, the mean waiting time was 0.39 weeks (2.73 days) and there were 182 redirections to other physicians per week, 32% better in wait time and the number of weekly redirections relative to the baseline. The optimized design reduces wait time and redirections by 40% compared to the baseline (Table 1). With regard to utilization, the optimized design and capacity-based design required fewer extra slots to be created than the base case and had a higher number of unfilled slots on average per week (spare capacity) (Table 2).

Increasing Panel Size

Wait times and redirections increase across all scenarios as panel size increases. In sensitivity analysis, the optimized design remained dominant over the base case up to an additional 2,000 patients (for the whole clinic). This remained valid with up to 3,000 additional patients if patient category was restricted to low-request patients (Table 3).

Adding Categories

Results for panel designs under this new patient classification are shown in Table 4. While the wait time and number of redirections are somewhat different, Optimal Design is still 40% and 36% better respectively in the two measures than the baseline.

DISCUSSION

Optimally redesigning panels has the potential to reduce wait times and maximize continuity. Our system matches physician capacity with historical demand from each category of patients better than the other strategies considered. Physicians who have less spare capacity in their schedules are given proportionally fewer patients from more appointment intensive categories and vice versa. The capacity-based design also performs quite well relative to the baseline for the same reason: physician capacities under this method are better matched with appointment demand than in the base case.

The optimized panels do this by increasing the effective capacity of primary care practices: as demand for appointments is better matched to capacity, many patients, who would otherwise wait to see their own provider, no longer need to wait. In addition, fewer follow-up appointments need to be made. Both these factors increase the number of available slots in future periods, with the result that more patients can be empanelled in the practice. Our model also produces similar improvements under an alternate patient classification system, suggesting that it is robust.

Computer-based models have been used in the past to study patient waiting time in outpatient settings. Factors considered include visit times,26–28, caseload of new or old patients,29, number of appointments a day,30, number of preceptors (in a teaching setting),31, or number of physicians or staff.32 These efforts are primarily target patient flow and reducing wait time in the clinic on the day of the appointment. Our model considers a clinic’s appointment system in relation to a physician’s panel size and case mix rather than patient flow and wait time in a clinic on a particular weekday. Wait time or timeliness in our model is the time from when the patient calls to when the appointment is secured; hence, the duration of the appointment on the day the patient sees the physician does not play a role.

Table 3. Effects of Increasing Panel Size

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Capacity-based</th>
<th>Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wait time</td>
<td>Redirections</td>
<td>Wait time</td>
</tr>
<tr>
<td>Current demand</td>
<td>Mean</td>
<td>0.572</td>
<td>266.65</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(0.570, 0.574)</td>
<td>(265.34, 267.96)</td>
</tr>
<tr>
<td>10% Higher demand</td>
<td>Mean</td>
<td>0.749</td>
<td>437.32</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(0.746, 0.751)</td>
<td>(435.05, 439.59)</td>
</tr>
</tbody>
</table>

Three panel designs (Baseline, Capacity-based, Optimal) are compared under current and 10% increased demand. CI stands for confidence interval. Wait Time is the average wait time in weeks for each patient. Redirections represents the average number of times patients requesting care saw a physician other than their own PCP in a week.

Table 4. Baseline, Capacity-based and Optimal Designs Under an Alternate Patient Classification

<table>
<thead>
<tr>
<th></th>
<th>Wait time</th>
<th>Redirections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline design</td>
<td>Mean</td>
<td>0.602</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(0.600, 0.604)</td>
</tr>
<tr>
<td>Capacity-based design</td>
<td>Mean</td>
<td>0.419</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(0.417, 0.421)</td>
</tr>
<tr>
<td>Optimal design</td>
<td>Mean</td>
<td>0.362</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(0.359, 0.364)</td>
</tr>
</tbody>
</table>

Alternate patient classification is based on age, gender, coronary artery disease (CAD), hypertension and depression. CI stands for confidence interval. Wait Time is the average wait time in weeks for each patient. Redirections represents the average number of times patients requesting care saw a physician other than their own PCP in a week.
Implications for Practice

Optimal panel designs obtained using our method would best be used, at least initially, as benchmarks or targets for real-world practices. It is not expected or desired that any real world clinic would necessarily reallocate panels abruptly, rather a more appropriate strategy would be to reallocate when the opportunities arise. For example, resident clinics have an opportunity to reallocate one third of their patients each year. Many primary clinic panels are dynamic, and patients enter and leave them all the time as people age, are diagnosed with new conditions, move out of area and many other reasons. A useful by-product of this constant state of flux is that it affords continuous opportunities to make incremental changes to patient panels without disrupting the visit patterns of patients who already have strong ties to their PCP, for example, leveraging patients who have yet to decide on a PCP, new patients and the turnover of existing patients. Patient surveys could be used to determine preferences and inclination towards change. In some cases, to minimize disruption, reassignment may simply be to another physician, whom the patient has seen almost as often as her own PCP, or to a physician within the same care team (if the care team consists of multiple physicians). At the very least, our model could provide pointers about how physicians in practice would benefit from enhanced care team support.

In making these recommended changes, the goal is to make steady improvements in timeliness and continuity wherever possible and continuously benchmarking against optimality. We envisage our model as a decision support system with a clinician-friendly interface that the office staff can use to test new panel allocations. This would enable immediate, structured feedback on the implications of changes on continuity and timeliness and therefore promote more informed choices. The model would be consulted on a weekly or monthly basis as panel adjustments are made.

Generalizability

Our study was conducted at an academic medical center with a substantial part-time work force. In this setting, because of research and education commitments, physician schedules change from week to week, affecting physician availability and hence timeliness and continuity. As a result our model had to tackle variation in physician supply up front. The model, however, also remains relevant for practices with a full-time work force where rather than physician supply the main drivers of supply-demand mismatch are panel size and case mix.

With appropriate modifications, our approach can be adapted to different scales. Specifically, it is applicable to the workings of a care team, to within a practice group, for a formal network of physicians affiliated with an HMO, PPO or hospital, to an informal network of physicians working within a shared geographic catchment area, for example the state of Massachusetts after the 2006 insurance reform. At the level of a care team, patient assignment among physicians, nurse practitioners and registered nurses needs to be carefully considered, while at the level of the network the appointment burden for different physicians needs to be balanced.

Limitations

Our study has important limitations. We do not consider individual patient and clinician preferences, which may play a role in how panels are formed. We also do not adjust for clinician practice style, which may impact the number of follow-up appointments. We do not account for operational adjustments that may occur on a daily basis. Physicians, for example, flex their immediate capacity by spending more or less time depending on whether the immediate demand is high or low; they may also use care team. We do not account for cancellations and no-shows—both of which are important components of the regular running of an office practice.

CONCLUSION

There is a large set of policies that may help address the problem of the primary care access shortage. These include alternative models of care25, well-designed care teams, payments for coordination of care, computer-based care and other tools to facilitate non-visit care and self-directed care for some patients. No one policy or intervention will solve the problem by itself.

We believe that increasing the effective capacity of physicians using systems engineering methods is not only an important part of the solution, but also a very cost-effective approach that should contribute to improved timeliness and continuity.

Acknowledgements: This study was partially funded by an internal grant at the Mayo Clinic and the Agency of Healthcare Research and Quality (AHRQ). We would like to thank Jason Egginton, Patricia Simonson, Sara Hobbs Kohrt and Liang Wang for help with preparing this manuscript. We would also like to thank anonymous reviewers whose comments have helped substantially improve the document.

Conflict of Interest: None disclosed.

Corresponding Author: Hari Balasubramanian, PhD; Department of Mechanical and Industrial Engineering, University of Massachusetts, 160 Governors Drive, Amherst, MA 01003, USA (e-mail: hibalasubramian@ecs.umass.edu).

REFERENCES

(46):211–18.


