Fraser Health Uses Mathematical Programming to Plan Its Inpatient Hospital Network

Pablo Santibáñez, Georgia Bekiou, Kenneth Yip,
Fraser Health Uses Mathematical Programming to Plan Its Inpatient Hospital Network

Pablo Santibáñez
British Columbia Cancer Agency, Vancouver, British Columbia V5Z 4E6, Canada, psantibanez@bccancer.bc.ca

Georgia Bekiou
Fraser Health Authority, Surrey, British Columbia V3R 4J7, Canada, georgia.bekiou@fraserhealth.ca

Kenneth Yip
Hospital Authority, Queen Mary Hospital, Pok Fu Lam, Hong Kong, China, ycm021@ha.org.hk

Fraser Health (FH), a British Columbia health authority that serves more than 1.5 million people, must increase its acute care capacity significantly over the next 15 years because of anticipated population growth and aging. The distribution of the projected capacity over each of FH’s 12 hospitals depends on the mix of clinical services to be provided at each site, a decision guided by population needs and clinical practices. We present a multiperiod mathematical programming model that we developed to provide options for configuring the system, specifically the location of clinical services and allocation of bed capacity across the hospitals. The decisions in the model are based on population access, critical mass standards, and clinical adjacencies. We describe its application in a long-term planning initiative that FH undertook. Extensive scenario analyses allowed administrators, clinicians, and planners to test multiple system configurations, gain a robust understanding of the trade-offs between these configurations, and formalize the planning process for acute care services.

Key words: health care: hospitals, service configuration; programming: optimization, integer, applications.

History: This paper was refereed. Published online in Articles in Advance March 4, 2009.
achieves clinical efficiencies, and serves population needs. In 2005, it launched the Acute Care Capacity Initiative (ACCI), a region-wide 18-month planning initiative to develop clinical service-delivery models customized to the needs of its population, both today and in the future. The main objectives were to understand current capacity pressures, prepare for the large expected surge in demand, and propose a long-term configuration plan for the network of hospitals. FH adopted a planning horizon of 15 years to address its need for a long-term strategy and the significant lead times required to plan and implement large-scale changes in health care. From its conception, ACCI was a data-driven, evidence-based process focused on population needs.

In this paper, we focus on one component of ACCI, the design of the network of hospitals through the location of clinical services and the allocation of bed capacity across sites. We present a service-siting and bed-allocation model used to develop configurations for the hospitals from 2005 to 2020. Through an iterative process with FH’s executive management, we developed a range of configurations for its consideration. FH decided on a preferred configuration from the final set of alternatives, devised a “directional plan” (i.e., a planning document outlining long-term strategies in terms of service delivery) for the acute care sector based on the recommended configuration for the hospitals, and proposed it to the provincial government.

**Configuration of FH’s Network of Hospitals**

The configuration of FH’s system of hospitals must accommodate projected demand. This comprises the location of clinical services across hospitals and the distribution of bed capacity by site and service.

In practice, little variation has occurred in the mix of services that each hospital provides; capacity adjustments have been the primary planning concern.
Although influenced by the perceived evolution of local population needs, current service configuration and utilization from previous years have been the main factors considered to identify service needs at each site. Capacity allocation has been based on the growth of the local community, adjusted for the role of the hospital in the system and the mix of services offered (e.g., because some specialized services are available only in a few hospitals, these hospitals serve the population of both their local community and a larger portion of the region).

When demand for a particular service increased beyond feasible expansions at current service locations, FH established a new service site in one of the other hospitals in the system. This site selection was also based on perceived population need and availability of supporting services at the possible hospitals. The site-selection discussion also included anecdotal information based on particular experiences, which were often exceptional cases. Although FH followed a series of logical rules, it had no formal process in place to determine the location of a service. Frequently, its decisions considered limited impact on the system as whole, and focused only on the service and hospital under analysis.

To develop the strategic plan for the next 15 years from a system perspective and effectively address the acute care needs of its residents in the future, FH decided to configure its services based on a data-driven, evidenced-based methodology in alignment with population needs and leading clinical practice. The approach needed to be sufficiently flexible to consider different configurations of the system (e.g., number of beds for any given hospital, degree of clinical specialization, options for future sites, and clarification of the hospital service role) yet rigorous enough to ensure clinical standards are met while providing good access to the population.

The elements of the problem are as follows:
- Demand (existing and projected future patient demand, grouped into 34 clinical services);
- Facility location constraints (12 acute care sites at present and options for future configurations);
- Geographical variation (multiple communities, grouped into 13 local health areas); and
- Multi-period decisions (three decision epochs in a 15-year planning horizon).

Considering these parameters, multiple configurations of the system are possible. It is impractical to consider and evaluate all configurations in a reasonable time frame without the aid of a planning tool. In addition, a list of clinical standards that FH must meet to site services increases the complexity of the problem. These additional principles, which FH clinical experts developed based on leading clinical practice, include the following:

1. Critical mass: Maintaining physician competence, care quality, and efficiency standards requires a sufficient population base. Critical mass measures usually take the form of physician-to-population ratios, minimum annual volumes per physician, average patient census per service location, or recommended unit size.

2. Clinical adjacencies: Some services must be provided in the same hospital facility to support the provision of care. In general, specialized, complex services require general services to be colocated in the same hospital.

3. Time-to-service standards: Service-specific response time or distance standards must be met. These are usually expressed in terms of a percentage of the patient population that must be able to access the service and receive care within a specified time limit.

Among all configurations that meet the previous considerations, those with the best performance are of interest. We seek and consider two goals to evaluate performance: provision of services close to the population and minimal disruptions (changes, service relocation) to the current system settings.

The Model

A key objective of this model is to provide decision makers with a series of viable sizing and siting configurations for their consideration. With the large number of possible hospital configurations, the requirement that every configuration must satisfy all clinical standards, and the consideration of multiple decision periods in the planning horizon, the problem becomes complicated and unmanageable without a systematic, quantitative approach. To resolve this, we developed a multiperiod mathematical programming siting model that allowed us to filter out impractical solutions (e.g., those that did not meet the clinical standards) and to
rank feasible configurations using predefined performance measures.

The nature of the problem makes an integer programming formulation suitable. Over a multiperiod planning horizon, there are a number of hospitals at which we can (1) site multiple clinical services, and (2) allocate capacity to serve the needs of their catchment populations. Such a model has a similar structure to a facility-location-allocation problem, extensively studied in the literature. Brandeau and Chu (1989) and ReVelle and Eiselt (2005) provide comprehensive overviews of location problems in general; Daskin and Dean (2004) describe specific location models applied to health care. In our case, the sitting decisions for clinical services correspond to the facility-location problem. The possible locations are restricted to the current hospitals and a few predetermined possible alternative and/or new sites. The allocation component corresponds to the capacity-allocation problem: the demand from customers in multiple locations must be assigned to the facilities. Similar to most location-allocation problems, demand is assumed to occur only at specified points (usually weighted-population centers), and the principal metric in the problem is a function of the distance to the facilities (in this case, drive times). Location problems usually consider multiple objectives, as we do in our problem. We combine a classic “pull” objective (Eiselt and Laporte 1995)—the minisum of transportation costs, which we represent by drive times, and the minimization of changes to the current system configuration (similar to fixed costs). Harsanyi (1975) provides arguments on the desirability of minisum objectives.

Doğrmeci (1977), Ruth (1981), and Stummer et al. (2004) consider similar decisions for the location of health-care services, although their examples relate to smaller instances (e.g., one service and decision period) or fictitious scenarios and numerical examples (e.g., theoretical applications of the models). Chu and Chu (2000) propose a modeling framework for service-location decisions, and through multiple lower and upper bounds for a variety of resources, constrain the allocation of bed capacity. Côté et al. (2007) consider the location of one specific service (traumatic brain-injury treatment units) within existing medical centers for one decision period. They recognize that the proposed model does not consider the effect of admission volume on the quality of care, and suggest minimum and maximum constraints to control the size of the units as a proxy to address this issue.

We know of no published studies that describe health-care location-allocation models that explicitly consider clinical standards. In addition, no published studies describe large-scale applications with multiple services, hospitals, and periods.

Our model includes two groups of decision variables. The first group (Y variables) is associated with the location part of the problem; it comprises 0-1 variables that determine if a service is placed in a hospital or not, in every decision period. The second set of variables (X variables) deals with the allocation component; it assigns demand for a service from each community to the hospitals, also in each decision period. The X variables are positive integer variables representing the number of patients assigned to a site. We can relax the integrality requirements for these variables because the demand is large enough so that fractions provide an adequate approximation.

These two groups of decision variables are linked: demand can be allocated to a site if and only if the site provides the service. The problem of configuring the hospital network is to decide on appropriate values for these variables such that the entire demand is satisfied in each decision epoch. We compute bed utilization using service-specific average length of stay (ALOS) estimates for each patient population and period. We compute the total capacity required at each hospital based on the mix of patients and services allocated to that hospital.

We note that this model deals with annual demand volumes; therefore, it does not explicitly address day-to-day variability. We account for short-term variation in bed utilization through average occupancy rates that are predetermined earlier in the planning process for each service and hospital size. We determine these parameters, which are an input to the model, based on acceptable service levels and efficient patient flow using historical utilization data.

The clinical practice constraints limit the allocation of services across hospitals. Specifically, the critical mass restrictions impose minimum service volumes at each location, thus affecting the allocation variables; the service adjacency constraints force the decision of
siting a service in a hospital, the location variables, to depend on the siting decision for other services. We developed a matrix representation of the clinical adjacency constraints (Figure 2), which we can recursively follow to construct self-sustainable clusters of services.

These clusters represent the building blocks of a hospital. Basic services are one-element clusters, whereas the largest clusters provide very specialized, tertiary-care services because they also have many other services available to support the safe provision of care. The model considers only mandatory adjacencies that are based on the provision of full inpatient services. This means that some components of a service, such as ambulatory clinics, can still be provided although the inpatient service is not located at the hospital.

Time-to-service standards force the service location to meet specific response times. Preliminary analyses based on population distribution and current hospital locations determined that the hospitals adhere to all time-to-service standards today—and will in the future, if the network has the required capacity allocation, regardless of the configuration of the system. Therefore, the model does not need an explicit constraint.

We impose additional constraints in the model to represent capacity and infrastructure limitations for each hospital. In some cases, the infrastructure surrounding a hospital significantly limits potential capacity expansions of the site and/or the types of services that the hospital can provide. We include other constraints to guarantee that the solution incorporates prior strategic decisions, such as predefined service locations, hospital roles, and service commitments.

Our problem has two goals for siting services—maximum population access and minimum disruption to the system. For the first objective-function component, we approximate closeness to the population by measuring total drive time by patients from all communities to all hospitals, for all periods and clinical services. Minimizing this measure provides better access for the population. For the second objective-function component, we define disruptions as changes in the service mix compared to the current configuration of each hospital. The sum of these changes over all hospitals and services corresponds to overall disruptions, which we wish to minimize.

The two objectives conflict to some extent. The population-access component seeks the minimum total drive time and therefore attempts to locate each service in as many sites as the clinical and infrastructure constraints permit. Current service distribution does not necessarily accomplish minimum drive time, requiring service relocations or disruptions to the current system. Conversely, the current service mix per site minimizes disruptions to the system but may possibly deteriorate population access if the total distance to services increases. The objective function in the model combines both components by using relative weights.

Cohon (1978) provides an excellent review of the different approaches to multiobjective programming, with particular references to public decision problems such as ours. He recommends solution-generating techniques for situations in which the analyst gives information (and solutions to the problem) to the
decision maker for consideration, on an iterative basis, as we did with FH’s executive management.

Our formulation of the disruption component in the objective function computes variations in service mix every period, counting the same changes multiple times. Disruptions earlier in the planning horizon count more than those delayed to later periods do. This is consistent with the concept that changes made sooner are more difficult to implement than those made later. If we allocate services to a hospital and then remove them (or vice versa), this formulation will not account for the second change because it will be the same as the original configuration. Depending on the conditions of the problem, this result might be undesirable. To avoid this situation, we can add supplementary constraints to force changes to remain for a number of periods, or until the end of the planning horizon. In our case, this was not an issue because the demand is nondecreasing over time, making service-location decisions permanent under normal conditions (e.g., nondecreasing capacity per site). Additionally, service-relocation costs could vary by service type: for example, certain services might require specialized equipment or infrastructure that also requires relocation. We can represent this in the objective function by assigning different weights to services and investigate the resulting trade-offs. This was important to the decision makers. This was a challenging process; configurations that were acceptable for some of the hospitals under a given situation were unsatisfactory under other circumstances. Our interpretation is that preferences are very difficult to state and model. Perhaps we need to incorporate additional objectives to the multiobjective function, although they are not obvious at this time.

Our model considers only inpatient services for siting decisions. We use the model results to derive volumes for other services or activities, such as day care and ambulatory programs; we apply service-to-activity ratios estimated from current practice and adjusted to reflect anticipated future service delivery.

The main inputs to the model are the set of hospitals (the network), the projected demand per service and community to be satisfied every period, and the clinical-practice constraints. The outputs of the model are service-siting maps (Figure 3) that specify, for every decision period, the hospitals that perform each service.

Based on the mix of services, we can identify each hospital’s role in the network—a rural hospital with only the most basic services; a local, community hospital; or a referral center with almost every possible service. The reports can also provide more detailed outputs, such as service-specific volumes per hospital and period, referral-pattern analysis (demand distribution over hospitals), performance measures (e.g., distances patients travel and changes to the initial system configuration), resource requirements,

\[
\begin{array}{ccccccc}
\text{Period} \ n & \text{Service 1} & \text{Service 2} & \text{Service 3} & \text{Service 4} & \text{Service 5} & \text{Service 6} & \text{Service 7} \\
\hline
\text{Hospital 1} & ✓ & ✓ & ✓ & & & & \\
\text{Hospital 2} & ✓ & ✓ & ✓ & ✓ & & & \\
\text{Hospital 3} & ✓ & ✓ & ✓ & & & & \\
\text{Hospital 4} & ✓ & ✓ & ✓ & & & & \\
\text{Hospital 5} & ✓ & ✓ & ✓ & ✓ & ✓ & ✓ & ✓ \\
\end{array}
\]

Figure 3: The model output specifies the mix of services to be provided at each hospital, for every period. Based on the resulting mix of services per hospital, their roles can be clarified, from rural hospitals providing only very basic services, to specialized referral centers providing almost the entire spectrum of services. In this sample output from the model for period \( n \), Hospitals 1 and 3 are identified as local hospitals given the limited services they provide, while Hospital 5 corresponds to a referral hospital given the full range of services.
and modeling parameters. Summary reports, such as the inpatient bed capacity per period for each hospital (Figure 4), provide an easy means of comparison between configurations.

We developed, implemented, and utilized this model as part of the ACCI, which is among the largest and most comprehensive health-care planning initiatives conducted in Canada. The clinicians and administrators involved in this initiative developed detailed acute care clinical-service plans that provided most of the information that the model required.

Implementation

We built the principal components of the model in an MS Access database that contains the definition of the system (e.g., hospitals, communities, and services) and the modeling parameters (e.g., demand, distances, travel times, and constraints). The system uses a form-based graphical user interface (GUI) that allows the planner to quickly develop and configure scenarios. We coded the mathematical model using the General Algebraic Modeling System (GAMS) and solved it using the CPLEX optimization software package, all of which we controlled from the user interface through Visual Basic for Applications (VBA) routines. We developed a reporting module that presents the model outputs in MS Excel spreadsheets, which the GUI generates automatically after reaching a solution. The reports include predefined summary tables and charts to allow easy comparison between configurations. In typical runs, a model instance has approximately 30,000 variables (1,300 of them are binary) and more than 9,000 constraints; the solver takes less than 15 minutes and achieves results within 1 percent of optimality.

Scenario Analysis and Results

We used this model to configure, generate, and evaluate more than 30 different configurations. We started with two basic configurations, each at one extreme of our possible objective function: (1) ideal service location based on population needs (closest to the population, no penalty for changes), and (2) least disruptions (fewest changes to current system, no access consideration).

We reported these initial configurations to FH’s executive management. With such an enormous amount of information for each configuration, we faced two challenges: presenting results in a concise manner, and comparing two or more solutions. We defined a set of key metrics to summarize, at a very high level, the performance of each configuration. These include average travel time per patient, number of immediate changes to current service configurations, bed capacity per hospital, and service self-sufficiency (percentage of residents receiving service within the local hospital) in every major region. These metrics became the first set of outputs that we showed for any configuration. More detailed service-mix maps and capacity-distribution reports provided additional information. We were able to perform preliminary comparisons between configurations using the summary measures. For a more in-depth understanding of the differences, we analyzed the service-location variables first, and then, if required, looked at demand-allocation patterns.

The response from the decision makers to this first set of configurations established the direction for the subsequent iterations. First, configurations based purely on population access (disruption component with null weight) implied numerous service relocations from current to ideal sites; FH management
deemed these impractical to implement. We thus increased the weight given to the disruption function. Second, the configuration obtained with the minimum number of changes to the system (no value on population access) added bed capacity across hospitals, similar to the current distribution. This entailed major, concurrent renovations at multiple hospitals to expand available capacity, while perpetuating population-access inequality at some currently underserved communities. This confirmed our belief that the ideal solution would be a trade-off between the two objectives.

In our next iteration, we developed a new set of configurations, varying the relative importance between the two siting objectives. The solutions continued to show significant capacity expansions at multiple hospitals. We had, however, expected this because the system had to grow significantly (especially in the beginning of the time horizon) to meet the forecasted demand. This originated a new stream of configurations to consolidate construction by focusing capacity growth in selected hospitals. The decision makers indicated that configurations that considered several sites undergoing expansion simultaneously would pose significant risk to safe-care delivery and that adding focused capacity would achieve efficiencies. The problem then became identifying configurations that would accomplish those efficiencies and understanding their impact on the system.

We next developed configurations that considered the addition of a new hospital to the system. We wanted to allow significant capacity expansion, but in a more focused manner. We estimated a geographic location for the new hospital site based on the community with the smallest self-sufficiency and preliminary studies on land availability. Results from the previous configurations were very useful for this decision: a new hospital could be beneficial in communities with high outflow and hospitals at maximum capacity.

We used this predefined location and tested different roles for the new facility. One configuration considered operating it as a community hospital. This would relieve the pressure on the local hospital and increase the numbers of patients served within their region of residence. As expected, results for this configuration showed a significant increase in self-sufficiency for the communities near the new hospital. This configuration also showed the impact on other sites in the network, particularly those with regional roles that, because of the reduction in community-level services, could now accommodate more regional-level demand and better serve other communities. This raised the question of whether a more specialized role for the new facility would help the system.

The next group of configurations tested the new hospital as a series of specialized centers providing services at multiple levels: lower-intensity services for the local community, medium-intensity for the local needs and those of nearby areas, and high-intensity for the whole region. We selected the different roles based on the services that accounted for a significant portion of the total demand or had major anticipated growth. Examples of such a role include a specialized surgical center, a maternal-child hospital, a comprehensive cardiac center, or an elder-resident-focused hospital. All these options would have an impact on the entire system because they might consolidate at regional-level components of services currently provided at existing sites, possibly altering their service mix. For example, it might be practical from a clinical perspective to sustain existing units in the specialized hospital and in other hospitals. In other cases, there was insufficient volume to make the new specialized facility a viable option.

We tested different variations in the role for the new site, from more comprehensive to less inclusive. In some configurations, the specialized role of the new hospital was partially covered by an existing facility that concentrated a significant portion of the region-wide services. In these cases, we tested configurations in which the existing facility expanded its specialized role (becoming a specialized center within the present infrastructure), while the new facility provided other services displaced from the existing hospital because of capacity constraints. We also evaluated different relative weights for the access and disruption components of the objective function. This generated different configurations for the hospitals based on the importance of distance to the population versus changes in service mix. We presented the results for each group of configurations to FH’s management for further refinement and consideration.

For reporting purposes, we present aggregated results for the three geographic regions that group...
the 13 FH communities. Because the scenarios that we analyzed mainly affected Regions 1 and 2, we report their performance measures, but do not show those of Region 3. To understand more clearly the effect that the new hospital would have in the local community, we report performance measures for the community in which we are considering locating the new hospital: Community 1 in Region 1.

To compare the results that we obtained for each configuration, one must understand the context. Numerical outputs, although useful, provide only partial information. Figure 5 shows performance measures relative to the current scenario (no service relocation or additions, only capacity adjustment over time to satisfy demand) for five illustrative configurations from those developed using the model; each considers a new hospital located close to Community 1.

The first conclusion that we draw from examining Figure 5 is that all configurations improve the self-sufficiency (proportion of the local demand treated in local facilities) of the local population, i.e., Community 1; the relative improvement ranges from 18 percent to 50 percent. Self-sufficiency measures for other regions in the system show limited or no improvement. In each configuration, the average drive time decreases between 11 percent and 19 percent.

To understand these results fully, we required additional information. We selected Community 1 self-sufficiency as a measure because this community currently has the largest outflow (lowest self-sufficiency), which is an indication of lack of capacity in the local hospitals. The other regions have an adequate self-sufficiency. The focus must be to improve the situation for the underserved community but not to impair that of the other regions. The degree of improvement depends on the services located at the new hospital, which vary according to its role. If its role is as a local hospital, then the majority of its capacity will be allocated to local residents, resulting in a significant increase in capacity. In contrast, if the new facility has a regional role as a specialized center with a system-wide scope, part of the capacity will be for the local residents; however, an important share will be for residents of other communities coming to the new center for specialized treatment not available elsewhere. If capacity is limited, regional volumes might push local demand for community-level services to hospitals outside the community to accommodate regional services. This results in deteriorated access for the local community accessing community-level services.

One particular area of FH’s region, a community with a large anticipated population increase, experienced significant capacity pressure. Therefore, we examined additional configurations for the hospitals, with special emphasis on this particular region, but within the context of the entire system. We used the model to support this process, and to provide alternative configurations for the service and capacity distribution across FH.

In both planning initiatives, we used multiple “what-if” scenarios to test alternative configurations. Sometimes, we had to provide additional insights into a particular solution. Our model allowed us to generate multiple configurations in a timely manner and gave us supporting information to back up service-location decisions. Testing, evaluating, and comparing all these configurations would have been impossible without a scenario-modeling tool.
The configurations developed with the help of our model have influenced and continue to influence bed-planning decisions and have informed capital and operating requirements. Recently, FH used our model as a guide in preparing for the immediate acute care component of its regional operating and capital budget from 2007 to 2010–2011, and for the longer-term period to 2020. In addition, the long-term directions that the model helped us to set have greatly influenced short-term decisions, such as expansion or reallocation of services.

The most evident benefit of this model has been the ability to develop, test, and evaluate multiple configurations in a reasonable amount of time; this allowed us to focus on exploring more options (rather than spending time in verifying the consistency of the solution), calculate the impact on each hospital, and determine if it satisfied the recommended clinical standards. The most important benefit, from a system-planning perspective, is the rigor and consistency used to create and analyze each possible configuration. For the first time at FH, we evaluated current state configurations and all the proposed changes to the system using the same methodology, based on the input from our own planning teams. The development of configurations, which (1) meet minimum clinical standards, (2) have clear objectives, and (3) are trackable in terms of the drivers for the decisions adopted, has helped to formalize FH’s planning and decision-making process.

Using this model has helped FH hospital administrators and clinicians to develop an increased understanding about the guiding principles for siting inpatient services; in particular, they now are able to consider the entire system rather than individual hospitals in isolation. The model enables system-wide network integration by considering all services for the entire population, including those services that the local hospital cannot provide but are considered part of a regional referral center.

Conclusions
Population growth and aging have been forecast to increase the demand for health care considerably in FH over the next 15 to 20 years—in particular, the demand for acute care beds. Although difficult, the allocation of a significant number of additional acute care beds into a network of hospitals also offers the opportunity to design a system to provide services that are accessible to the population and meet clinical standards. The development of this siting and allocation model allowed FH to test multiple configurations for its hospitals and to understand the trade-offs between various options. The establishment of this planning methodology and its associated performance measures helped to formalize the quantitative elements in the decision process. Population needs and evidence inform siting decisions; results can be tracked and supported with data and the clinical siting principles.

The model incorporates siting rules based on leading clinical practice, a feature not available in any health-care location-allocation models in the literature. The siting rules include clinical adjacencies and unit size to ensure safe provision of care. The inclusion of such rules has been fundamental in representing the problem and in getting agreement among the stakeholders.

In this model, we assume that the demand can be assigned to hospitals. In reality, patients have some choice of the hospital in which they receive care; their perception of the quality of care at each site influences their decisions. Furthermore, utilization patterns are subject to referrals from individual physicians—a practice difficult to modify from a centralized perspective. Nevertheless, FH can take several actions to influence where patients receive care. One strategy is the allocation of capacity at the hospitals with high demand (available capacity is a factor in referrals by physicians). Second, FH can inform the public and the practitioners of predefined referral patterns. In addition, FH can transfer patients within its network once they are users of the system, adjusting utilization patterns. These actions will induce some change in the system, but other factors outside FH’s control could determine actual utilization. The objective of demand allocation in our model was not to identify exactly which patients receive care at which hospitals, but to distribute capacity consistent with population needs.

The health-care system will continue to increase in complexity. The use of models, such as the one
we developed and used in FH, allows decision makers, clinicians, and planners to focus on identifying and evaluating more configurations and what-if scenarios. As the health-care system evolves (e.g., through changes in technology, clinical practice, and population demographics), changes in the service mix and capacity allocation at hospitals will be required. The use of this model ensures consistency in evaluating these changes.

The model is applicable to similar configuration problems in other jurisdictions, situations with shorter planning horizons, and other health sectors that require the allocation of resources over multiple locations.

At the time of this writing, the British Columbia provincial government is considering the proposed configuration plan for FH hospitals, in particular, its capital-funding implications. It might be premature to call this study a “success” because a final decision has yet to be made. Nonetheless, we were successful in that the configurations we analyzed in this planning initiative were useful and relevant to executive management in developing a hospital configuration plan for FH.

Appendix. Mathematical Programming Formulation

Sets

c: communities (demand points).
h: hospitals.
s: clinical services.
t: periods in the planning horizon.
ADJ(s): set of clinical services that need to be adjacent to service s.

Data Parameters

$D^t_{cs}$: projected patient demand for community c and clinical service s on period t.

$CM_s$: critical mass requirement (minimum volume of patients) for service s at any given hospital.

$LBed_s$: lower bound for total number of beds (all services) to be allocated at hospital h every period.

$UBed_h$: upper bound for total number of beds (all services) to be allocated at hospital h every period.

$\tau_{ch}^t$: drive time from community c to hospital h.

$\alpha$: relative weight for the two components in the objective function ($0 \leq \alpha \leq 1$).

$I_{hs} = \begin{cases} 1 & \text{if service } s \text{ is currently available in hospital } h, \\ 0 & \text{otherwise.} \end{cases}$

Decision Variables

$Y^t_{hs} = \begin{cases} 1 & \text{if clinical service } s \text{ is allocated to hospital } h \text{ on period } t, \\ 0 & \text{otherwise.} \end{cases}$

$X^t_{chs} = \text{number of patients from community } c \text{ assigned to hospital } h \text{ for clinical service } s \text{ on period } t.$

With $\tau_{ch}$, the drive time from community c to hospital h, we approximate closeness to the population by the total drive time spent by patients from all communities to all hospitals ($\sum_c \sum_h \tau_{ch} X^t_{chs}$ for service s and period t), for all periods and clinical services. Using the parameter $I_{hs}$ to identify whether service s is currently available at hospital h, the difference between the service-allocation variables $Y^t_{hs}$ and this indicator ($|I_{hs} - Y^t_{hs}|$ for service s at hospital h in period t) represents service-mix disruptions in the system. With $\alpha$ specifying the relative weights of the two components in the objective function, the weighted sum of our proxies for closeness to the population and disruption to the system is the objective function to minimize in our model.

The complete formulation of the mathematical programming problem is as follows:

$$\begin{array}{l}
\text{Minimize} \quad \alpha \left( \sum_{h} \sum_{s} \sum_{t} |I_{hs} - Y^t_{hs}| \right) \\
+ (1 - \alpha) \left( \sum_{c} \sum_{h} \sum_{s} \sum_{t} \tau_{ch} X^t_{chs} \right) \\
\text{subject to} \quad \sum_{c} X^t_{chs} \leq Y^t_{hs} \cdot M^{ops} \quad \forall h, s, t, \\
\text{where} \quad M^{ops} = \sum_{c} D^t_{cs} \text{ is an upper bound based on the demand for the service} \\
\sum_{h} X^t_{chs} = D^t_{cs} \quad \forall c, s, t, 
\end{array}$$

(1)
In the above model, constraint (1) links the two decision variables of the problem by allowing patients to be allocated only at sites where the model is selecting the service to be provided while forcing the others to have no demand assigned.

Constraint (2) ensures that for each community, all the demand for every service is allocated in the hospitals, during every period. Constraints (3) and (4) represent clinical standards. Constraint (3) imposes critical mass requirements by forcing minimum patient volumes per service location; constraint (4) forces the colocating of services that are listed in the adjacency matrix.

Constraint (5) sets lower and upper bounds on the total number of beds per hospital, where \( ALOS_{cs}^t \) represents the average length of stay for each service on a community- and period-specific basis, \( \delta_t \) is the duration in days of period \( t \), and \( \rho_{hs} \) is defined as the planning occupancy rate of the service at a given hospital. \( (X'_{chs} \cdot ALOS_{cs}^t / \rho_{hs}) \) is the number of patient bed days allocated to the hospital, with the length of stay adjusted for the desired occupancy rate. The quotient between the sum of this term over all services and communities and the duration of the period is the total average number of beds that will be required at a given hospital. This is limited by infrastructure (upper limit) and possibly policy decisions, such as maintaining a minimum number of beds in the site (lower limit).

Constraint (6) defines each \( Y_{hs}^t \) variable to be binary, and constraint (7) sets \( X'_{chs} \) to be positive. Although assigning fractions of patients to hospitals is unrealistic, the integrality constraint can be relaxed considering that (1) the demand for services is reasonably large; (2) there are minimum volume constraints that prevent the allocation of small volumes, where the problem can be more relevant; (3) the demand forecast has an error range larger than a few patients per service; and (4) these are annual patient volumes to estimate high-level capacity needs.

Additional constraints used to shape the configuration being modeled include limiting the number of sites per service to test the effect of distributed or consolidated service-delivery models \( (LBeds_h \leq \sum_{s} X'_{chs} \cdot ALOS_{cs}^t / \rho_{hs} \leq UBeds_h) \) for every service and period of time, forcing services to stay at a hospital once sited there \( (Y_{hs}^t \geq Y_{hs}^{t-1}) \) as a representation of consistency in service mix over time, and critical mass constraints expressed in bed days (minimum unit size) instead of volume of patients \( (\sum_c (X'_{chs} \cdot ALOS_{cs}^t) \geq CM_s \cdot Y_{hs}^t) \).

Acknowledgments

We thank the more than 385 physicians and administrators from Fraser Health who collectively developed the majority of the clinical-planning parameters used in the service-configuration model presented in this article, and provided invaluable input to turn the early solutions of our model into realistic configurations that make sense and incorporate leading practice principles. We are also grateful to the other members of the ACCI project team, Darlene Hope-Ross, Irene Chanin, Yurik Sandino, and Victoria Ostler, for their contribution throughout the entire project. We also thank the anonymous referees for their very helpful and constructive feedback and numerous recommendations to improve the quality of this paper. All three authors were affiliated with Fraser Health Authority at the time of the study.

References


Patricia Petryshen, Executive VP, Acute Programs, Fraser Health Authority, 300-10344 152A Street, Surrey, British Columbia, Canada V3R 7P8, writes: “This is to verify that the mathematical model described in the paper ‘Fraser Health Uses Mathematical Programming to Plan Its Inpatient Hospital Network’ written by Pablo Santibáñez, Georgia Bekiou, and Kenneth Yip has indeed been developed and used, and that it has been very valuable to our organization.

“Although it is too early to identify quantifiable benefits, as this model is intended for performing long-term planning, the qualitative benefits that it brings to the organization are substantial.

“During our recent formal long-term planning process, the Acute Care Capacity Initiative (ACCI), the model allowed administrators, clinicians, and planners to evaluate multiple service configuration scenarios. This would have not been possible without a model, considering the level of complexity in planning for our hospital network; especially in building a system where access and all the necessary clinical safety criteria are satisfied.

“This model has helped us to formalize our acute strategic planning process under a data-driven, evidence-based approach. The use of such a methodology for planning our hospitals into the future is a pioneering approach in the Canadian health-care industry.

“In conclusion, I am pleased to advise that we support this model and will benefit from its application after the conclusion of the ACCI in our various regional long/mid-term planning initiatives.”