

Review of the Canadian Statement on Wait Times in Obstetrics and Gynaecology: a SOGC Policy statement, published in JOGC, No. 204, March 2008.

Wait time benchmarks were recognized as a government priority in 2004. The Wait Time Alliance for Timely Access to Healthcare was established and the federal, provincial and territorial governments agreed to establish benchmarks for wait times in five priority areas (cancer, cardiac surgery, care, diagnostic imaging, joint replacement, and sight restoration). In April 2007, the WTA was expanded to include emergency care, psychiatry, gastroenterology, anaesthesiology and plastic surgery. The authors state that “here are currently no initiatives to manage wait times in the fields of obstetrics and gynaecology and benchmarks for appropriate wait time are lacking.” The objective of the document reviewed herein was to “bring forward expertise-based recommendations for consultation, procedures and surgery within obstetrics and gynaecology.”

To achieve these goals the SOGC recommends:

- standards or thresholds for delays for consultation, for investigation (procedure) and for surgery in both obstetrics and gynaecology,
- the establishment and support of an appropriate triage process within the framework of the wait list benchmarking process,
- that “to ensure quality of care, the SOGC recommends that 90% of these services and procedures be available to women within the time recommended.”

Acknowledging that “Inappropriately long waits may adversely affect (sic) health outcomes, and, in some cases, result in mortality”, the authors proceed to propose benchmarks that are “performance goals that reflect a broad consensus on **medically reasonable** wait times for health services delivered to patients.” SOGC endorsed the ten principles proposed by WTA for the development of wait time benchmarks. These principles include the provision that “targets to reduce wait times must be sustainable”, that benchmarks “must not be achieved at the expense of reduced access to other health care services” and “must be implemented with appropriate guidelines and prioritization tools that are fair, equitable and transparent to the patient”.

To develop the benchmarks the authors surveyed the gamut of clinical care areas of practice within those fields, divided the clinical workloads into three sectors: consultation, investigation (test) and surgery and then developed independent wait time benchmarks for each health care service within these three categories for both obstetrics and for gynaecology. The thresholds were set at a “level of severity beyond which best available clinical evidence and clinical consensus indicates that the patients health is likely to be adversely effected.” Six thresholds were set in obstetrics and seven in gynecology. In obstetrics time limits for each of the three categories were set at within 30 minutes, between 30 and 60 minutes, between 1 and 2 hours, within 1 week, within 4 weeks and as required by standards of care. In gynaecology, threshold limits were set at within 2, 3, 46, 8, 12 and 26 weeks.

Analysis

This approach suffers from serious and fundamental flaws. These flaws are not due to a belief that some clinical conditions are different or more time sensitive than others or due to the specific division of that work- the priority categories - but the flaws are operational in nature:

- Total system or department capacity has to equal total system or department demand. Without this balance, and in particular, if demand in any of the distinct demand streams discussed (consultation, investigation, and procedure) exceeds supply, then setting a wait time threshold or standard is meaningless, since that mismatch will create an inevitable delay.
- Prioritizing demand into various sub-streams with standardized waiting time goals or thresholds will require the use of sophisticated triage. This triage resource will use up capacity to sort the work that could have been used to do the work. As a consequence, demand will continue and supply will be diluted due to resource required for triage, for training and updating. Thus, supply needs to exceed demand in order to make this proposal work.
- Prioritization will inevitably result in mistake and error. This error will occur at the point of referral (inappropriate or inadequate referral) as well as at the level of triage. Even “perfect” information is useless in a changing clinical environment. Error leads towards a desire for even more prioritization, more triage and more inquiry, which will create even more channels, standards and lines—which in turn will create a higher likelihood of delay.
- Prioritization will increase the likelihood of no-shows and unused capacity in the lines with longer thresholds. Some of these patients will balk or renege and quit the line, while others will cut in line (or their Primary Care referring doctor will help them cut in line) and they’ll find entry into the system through a shorter portal (through the on-call function or into the hospital). These activities will increase the chaos, rework, phone calls and inquiries in the lines, and increase the no-shows.
- Multiple priorities and entry criteria will lead to exaggeration and fabrication of symptoms on the part of the referring providers in order to get their patients more quickly to the front of the line. This increases the re-work and redundancy in the referral system and increases the likelihood that referring providers will call the specialists personally to advocate for their “exceptions”.
- The authors fail to recognize the need for system measurement, both in the overall level and within the individual streams. Prioritizing the workload (demand) within each of the different streams based on clinical condition will create multiple channels of work. Each of those channels of work will demonstrate either demand or supply variation. These variations will create temporary mismatches in supply and demand in all the various channels. This variation in demand or supply will be extremely difficult to manage. The greatest effect of variation will be on the shorter, more critical, time-sensitive lines:
 - In gynecology, for example, and obstetrics faces a similar situation, patients with 26, 12, 8, 6, 4 and 3-week thresholds will be scheduled at the margin or limit of those thresholds. Although demand into the 2week line is predictable, it is not fixed and it will vary within a predictable range. Then, as variation occurs into the 2-week line, there is little flexible capacity to manage demand that exceeds the amount of supply or capacity allocated. The practices then have to either overbook into the 2-week line or steal capacity from one of the already pre-scheduled lines by overbooking into those lines or canceling from those lines. Stealing supply or capacity from another line will cause that line to exceed its threshold and “break” one of the WTA principles (solve one problem at the expense of another). As an additional consequence, more staff resources are needed, more triage and priority is needed, and the likelihood of delay for those patients deemed to be the most critical actually increases. Ironically, then, the patients with the highest likelihood of having a delay past their threshold are not those patients with the lowest priority, but those patients with the highest priority. With multiple queues accompanied by set thresholds for each queue, demand variation will guarantee delays in the line with the highest priority. In queuing theory, if supply is fixed (and in these systems supply is not

- fixed. Supply is highly variable.) and demand varies with common natural demand variation, trying to balance more than 2 lines with set, pre-determined thresholds is impossible. It is operationally thus impossible to achieve the goals of this proposal.
- The expressed goal of achieving 90% compliance makes no sense. First, this goal is difficult to measure. Second, if this goal means that 10% of the time, the practice, department or individual can “slip over the line”, slipping over the end of the line has severe system consequences in terms of both clinical risk and patient satisfaction for the line measured and for all the other competing lines. The 10% fudge factor is often introduced as an indirect way to deal with variation. It is far more effective to deal with variation in other ways.

Wait time issues are primarily operational, not clinical issues. “Medically reasonable” prioritization sounds reasonable. After all, some patients are sicker than others and deserve to be seen sooner. However, prioritization as such is built on fantasy- the fantasy that demand equals supply, that all the priority queues will contain exactly the same amount of workload as the previously allocated supply and that there will be no variation. This fantasy hopes that not only demand will exhibit no or little variation but assumes that there is no supply variation. Most of the clinical conditions described are managed within the office or scheduled procedure setting. In any given time frame- a day, a week or a month- the amount of supply variability is staggering. Some of the proposed wait time threshold work, particularly in obstetrics, will be managed by supply consistently allocated each day, often as an on call function. (The delay thresholds described in minutes or hours will always have supply allocated in anticipation of such needs. Demand will vary but at least supply is fixed) On the other hand, for some of the longer thresholds, (“as required by standards of care in obstetrics and 26 weeks in gynaecology), will have lower priority and much more high variability in supply. Hence, the variation in either demand or supply will have the greatest effects on those with the shortest or the longest delay thresholds respectively. The patients with the highest perceived risk will have the highest likelihood of not being seen within goal. This proposed system is designed to fail, is designed never to achieve its proposed outcomes. As such, it can be expected that practices and providers will find methods to accommodate, to work around and to cloud measures of system performance to avoid judgment and scrutiny since they simply cannot achieve these goals.

Alternative Recommendations

- Measure total system capacity and demand across all demand streams (all types of work), and achieve a balance between that demand and the total capacity. Demand streams- distinctly different work- includes: office, procedure, surgery-operating room, surgical center, hospital, the on call function etc
- Measure the demand, the supply, the activity (what is done) as well as the delay for each individual demand stream and achieve a balance
- Measure the demand, the supply and the activity as well as the delay within each demand stream. For example, the office demand stream contains new work and return work. The total of that demand must equal the supply or capacity allocated to that work or an ever-expanding delay will ensue.
- Minimize the number of lines and queues and priorities
- Reduce the wait time goal for everything to below 2 weeks in gynaecology and within 1 week in obstetrics (except for the delay thresholds measured in hours or less which have to be managed by an on call function). This allows load-leveling of all the various queues, and far more flexibility.
- Flex the capacity to keep up with demand. This can be done with pooling of the referrals to avoid individual mismatches and also can buffer the effect of supply absence, changing the ratio of office new appointments to return appointments, and making sure there’s enough provider new patient capacity in order to keep up with predicted and measured demand.

While the SOGC is to be commended for addressing the overall issue of delays, this approach, based solely on clinical care priority, is operationally dangerous and, quite frankly, unable to achieve its stated goals. It is possible though to achieve the wait time goals but this requires measuring and balancing overall demand and supply for all conditions and addressing demand or supply variation by minimizing the channels of priority.