Appointment Types in Primary and Specialty Care

Primary Care (PC)

In PC, if we can get the wait time to zero, then all appointment types that we used to call "urgent", "acute", "same day", non urgent-routine, and the pre-booked returns can be managed with a single appointment type that drives the work to the linked panel provider. So the sorting is done by provider not by clinical urgency or condition. These appointment types include both external (outside of us) and internal (return) visits. Sometimes these visits are called "short" appointments. "Long" appointments are a sub-set of this first appointment type and include visits that are expected to take more time.

A second appointment type (this is a third appointment type if "long" is distinct) can be used to share the work from the absent providers across the practice, to keep that work from going disproportionately to one provider.

Specialty Care (SC)

In SC, the primary goal is to get new patients in without a wait. Because there is a delay, new patient appointment types have to be distinguished from return visits. The external demand is constituted primarily by new patients, but also contains a smaller number of established patients with acute recurring problems. The patients asked to return constitute the internal demand. "Long" appointments are generally new patient visits while "short" are the returns (internal demand).

Thus, "short" and "long" have different meanings in PC and SC. In PC, the "long" is a sub-set of external, and in SC, this is the bulk of the external. In PC, "short" covers most visits (urgent, non urgent, return, etc.), while in SC, "short" includes all returns (internal demand).

In PC, the focus for reducing wait times and the measurement of TNA is most valuable in short appointments whereas in SC the most important TNA is the delay for long appointments.

Analogy

The analogy of the movie line is often used to describe the mechanisms of queuing for appointment types. If the total or average demand volume equals the total or average staff handling volume over a specified timeframe, then if a movie theatre has 4 lines (1 for each movie playing), the variation in any one line either going up or down in demand volume can cause delays in a high demand line or unused staff capacity in any of the low demand lines. On the other hand, if all staff could take tickets for all movies (as is the common case) then volume variation is load levelled across all 4 lines and the customer in line wait times are significantly reduced.

In PC, this analogy applies for appointment types at the individual provider level. The movie theatre represents each individual provider and the lines are the various appointment types directed to that provider. In a sense, the linkage between patient and provider is pre-determined—that is, determined in advance of the appointment. Patients know in advance the identity of their provider and the lines or reduction in lines designate the types of work directed to that provider.

In SC, while the analogy still applies, it applies in a slightly different context. The movie theater is the "department" or collection of like and interchangeable providers. The lines are designed for new patients. In this analogy, some of the staff can leave for breaks and not be available to support the lines for extended periods of time. This represents the SC provider who is "busy" with other competing work, supporting, in a sense, an entirely different theatre. The remaining staff can then take the next available customer in line. In this way, no lines are left unsupported with the consequential delays due to absence of staff.

The mechanism and analogy of the queues and theatres still applies but there is a difference in application. In PC, the analogy applies to division of work at the PCP. The workload that goes to the PCP results in less delays if that workload can be managed universally, with as little divisions (appointment types) as possible. In SC, the analogy applies to specific workload (new patients) at the department but not the individual level. If that workload can be leveled so that any new patient is directed to the first available staff (representing a new patient appointment), then waiting times for any absent staff person can be avoided. This difference in analogy can be accomplished because new patient linkage with a specific SC provider does not have to be pre-determined. In practice, this approach has been called "pooling of referrals" which allows referrals for new patients to go to the department to the first available new patient appointment slot rather than be referred in a pre-determined way to a specific provider. In the analogy, referring to a specific staff person who may be absent will create delays for that specific staff person, but if referrals for new patients can be managed by any staff person, the delays are reduced. In the practical world, this can work transparently. Patients do not arrive at the theatre not knowing who will serve them but due to information technology management knows in advance which staff will be present or absent. To accomplish this though, SC have to allow pooled referrals and eliminate to the greatest extent possible referrals to specific providers who may or may not have open new patient slots for a period of time.