



## Measurement of Demand as "Bodies" vs. Time

One of the key measures of our performance measurement triad is the measure of demand, supply and activity. This is a granular measure of how well we use supply (capacity) to meet demand. The demand is the workload, the supply is the amount of worker poised to meet that demand, and the activity is the measure of how well we used the supply to meet that demand. The relationships between these three measures are critical for analyzing system performance. In order to see and analyze these relationships we need to collect and calculate these three measures in the same units.

Groups will commonly calculate demand as number of appointments made ("bodies" or number of patients) but calculate supply as time (number of hours worked). Using different units to measure demand and supply makes comparison difficult. For groups just getting started in this work, I recommend calculating demand and supply as bodies for the following reasons:

- This is the simplest and easiest way to measure. The unit of measure (bodies/number of patients) is easy to understand, count and measure.
- The unit of measure (bodies/number of patients) is objective.
- Activity is commonly measured as number of patients seen. All three components (demand, supply and activity) are therefore recorded in the same unit.
- The comparisons make intuitive sense: We had X number of bodies requesting appointments, so many slots designated to receive those bodies and we saw so many bodies/patients.

The problem is that because each body/patient is variable in the amount of work effort required for resolution, measurement by bodies does not always reflect that required work effort for each visit. This tension that bodies do not equal work effort is seen more commonly in salaried or pay-by-the-visit systems and less commonly in classic fee for service (do more, get more) since the classic fee for service practices tend to measure "productivity" as revenue (RVU) and not as "units processed" (bodies or time) which is the most common measure of "productivity" in salaried or pay-by-the-visit systems. In salaried or pay-by-the-visit systems, since the "pay" is the same, providers and their managers are interested in making sure that the work effort is the same, to balance against the same salary. Measurement by time is an effort designed to account for this variation. However, there are challenges with measuring by time:

- The demand measure is subjective: this body = how many units of time? The schedulers have to make this decision.
- This decision can be influenced, particularly by providers.

- This decision is made prospectively, in front of the visit and is commonly inaccurate. The decision is made with the worst case scenario in mind and is often based on fear of just how much time this will take. The time decision is a "guess."
- There is a tendency for staff and providers to want to "inflate" or over-estimate the demand. This gets reflected in the subjective decision of demand: how much time is this body/patient worth? The decision may have multiple self-serving elements (help me reach my time quota with less bodies of workload, help validate my victimhood, help me have less work).
- The supply units have to be flexible to meet the variable decisions around demand or a delay will ensue. If there are a limited number of "long" or "longer" appointments on the schedule template and demand for those exceeds the supply on the template then these are "booked" into the future with a delay. This can be mitigated, of course, by having smaller units of time that can be merged into longer units to accommodate that increased burden of demand. Having smaller units of supply on the schedule template makes the calculation of supply a challenge. We could, however, just "count" supply as, say, four hours of supply for each half day session and then compare to the demand. Supply then has a limit whereas counting supply as body slots has some degree of flexibility. I wonder then with a fixed supply, if we wouldn't run into the problems described in Erlang's formula where a variable demand meets an inflexible supply.
- When supply is time in hours and demand flexes according to a subjective decision about how much time each request is "worth" we commonly end up with not enough activity to "pay the rent." The demand is inflated and the supply absorbs that demand but the activity is low.
- Due to the tendency to "inflate" the burden of the demand due to fear, a prospective measurement approach and responding to the worst case, we commonly see the demand appears to be "worth," for example, 10 hours but the providers actually finish the work in 8 hours.
- Some groups avoid the subjectivity of a prospective demand measure by counting demand as what was done, or how much time was taken. However, then we are measuring activity as demand. Measuring activity as demand tends to perpetuate some self-serving behaviors and blurs the value of seeing the relationships between demand, supply and activity.

How to reconcile the drawbacks of measuring in these two very different ways:

- Start with measurement by bodies/patients and measure all demand, supply and activity in that way.
- Consider a measure like RVU (revenue) either as a sole unit of measure or as a balancing measure that converts bodies into various levels of work required. RVU is a retrospective measure and, as such, eliminates the subjectivity and self-serving behaviors seen with prospective measures.
- If you decide to move toward measurement of DSA by time:
  - Develop objective criteria for which demand requests need more time, and how much time.
  - Make the criteria of differentiation simple and objective.
  - Limit the number of visits that can result in long or longer time unit decisions.

- Eliminate the opportunity to influence the choice.
- Allow the supply to flex past a rigid schedule time.
- Continuously monitor the time recorded as demand and the time actually worked to manage the workload.

At its core, this issue reflects a concern about how to account for the known variation in demand within each visit. The desire to reconcile this variation most commonly occurs in systems where the salary is fixed or the practice is paid the same for each visit. We can attempt to account for this variation in demand by counting demand as time but this is a prospective measure and, as such, is subjective. It is a measure of what might happen. We can, on the other hand, account for variable demand by a retrospective measure, such as RVU. This is a measure of what did happen. Some groups are just not prepared to measure RVU and for some groups with capitation or "alternative reimbursement plans," RVU is not a common or relevant measure.

All in all, I like the measure of demand, supply and activity as bodies and not by time. I recognize that all bodies do not require the same work effort and there is variation. At the same time, with enough visits, these variabilities in work effort required will balance out. Some groups recognize this "balance out" and allow a daily variability in visits but set a long term (yearly) expectation for visits as a measure of "productivity." This approach accounts for daily variability by flexing activity but maintains an overall blended expectation for productivity as visits. There is a recognition that while there may be a daily demand variability in required work effort, that this variability both rises and falls. In addition, if there is more effort required then we ought to look first at what is the difference in that requirement and how much of that difference we can eliminate with a better care team workload analysis and change. Often, in my experience, what is perceived as a required increase in work effort really is not provider work at all but work that could be done by another team member in the context of an activated team.