Health Economics

Dr Pratap C Mohanty

Department of Humanities and Social Sciences,

Indian Institute of Technology Roorkee

Week – 03

Lecture 12- Physicians as Health Provider-II

Welcome friends. Once again, regarding our NPTEL MOOC module on Health Economics, we are in our third week, or unit number 3, explaining the supply side of health or supply in healthcare. We have already started in the previous lecture on physicians as healthcare providers or health providers. This is our lecture number 2nd sequence of this particular lecture. We discuss supplier-determined, primary, and supplier-induced demand (SID).

Trends in the supply of physicians also discussed supply-induced demand from increasing physician density and conditions facilitating demand inducements. The last couple of points will also be re-emphasized in our lecture. Again, we focus on the supply-induced demand through the physicians. This is as follows.

We will talk about the physician's behavior through a model and empirically examine the behavior and how it is optimal for the supply side. There are some forms of discrimination made by physicians. One type of discrimination is racial discrimination, or maybe through efficiency and non-efficiency-based discrimination. We will emphasize. Last but not least, the aspect covered in this lecture is that the labor market and wages are responsible for determining healthcare supply through physicians.

Once again, counting the assumptions, you can follow our previous lecture. For your knowledge, p, we are taking it as price, and a is largely the physician, which is considered to be identical. Hence, they are considered as identical physicians. The n is the inhabitants or the persons or the patients. Delta stands for a upon n.

We count this as physician density or physician-population ratio. So, this is the one we will use the most in our modeling. Our model is on physician behavior. We start with no demand inducement, and if there is an inducement. If no inducement is as simple as that, that is called primary demand for physicians.

It is precisely the cases or number of cases times the population divided by the identical physicians. So, a by n is precisely called the delta (δ). So, M upon δ is called primary demand for physicians. When you use inducement, we have already discussed how inducements are made in healthcare, and that is as we have seen in a number of cases in the Indian context, particularly since the supply side is quite less than the demand side. Hence, supply inducement is expected to be very high, so supply hospitals or healthcare units provoke patients in various ways to increase their demand.

Hence a component we are just adding which is from the primary demand pattern M upon delta plus s, s stands for the unit of inducements. So, M, we have already discussed what is called demand for physician services. So, we will be discussing all these in our model. And out of the total time, the time as a factor we are considering as well in the model and we are also referring to the original author. There are some minor cosmetic changes to the author.

We will also cite author's work at the end. And in this case t proportion of time that is utilized for healthcare and to induce the demand, we are saying that the optimum time proportion is 1 and out of that at minimum of h time or h units of time is required for healthcare. And that is precisely a function of the induce content or the units and the δ. So, delta we already mentioned is they are the density of the physicians. We have already mentioned this physician density or physician-population ratio.

So, time constant we have emphasized. We discussed disposable income and how it inspires the physicians or the health units to influence the demand. So, the disposable income of the physicians is largely a function of the revenue, which is basically p times t. Here p is the price, and t is the time, practice expenses, taxes, etc. Lastly, in the last lecture, I discussed that y is actually other p and t practice expenses, and taxes are already part of the revenue pt.

Hence, the simple function we mention is that y equals y of p t. So, this is considered a concave function with its feature as the first-order derivative being positive and the second-order derivative negative. Hence, disposable income is an increasing and concave function of physicians' consumption. Here, the utility function is defined, and the function is defined in terms of disposable income, time, and induced demand. The physician's utility function is strictly concave.

Disposable income of physician, y = revenue (pt) – practice expenses – taxes

Let

- → Practice expenses are fixed share of revenue
- → Taxes are progressive

Then,

$$y = y(pt) \qquad \text{with } y' > 0 \text{ and } y'' < 0$$

Disposable income is an increasing and concave function of physician's consumption.

It depends positively on consumption and negatively on working hours or time and demand inducements. There are three possible conditions. One is its first order derivative will be positive and second order negative that, is precisely called concave function and with respect to that is with respect to income, first-order derivative with respect to income and what happens in terms of t. This is because the first one is in terms of y, the second one is not t, and the third is in terms of the induced demand units. So, there are possibilities with respect to the utility: when it increases, the utility declines, and in terms of induced demand, it also declines.

There are also cross possibilities, complementary goods case when y and t complementing each other, consumption and leisure are complemented. Similarly, income and induced demand that is also complement each other. We also tried to find their cross-second-order derivatives. In that case, I observed that the second-order derivative for y and t is considered non-positive and others like y and s is also non-positive. Whereas in case of s and t that is, workload time has no effect, it is considered to be 0.

The second-order derivative here is considered to be 0, which means that workload time has no effect on physicians' attitudes towards professional ethics. So, time has hardly any effects in terms of professional ethics. Hence, the change is considered to be 0. So, we are just highlighting this just to clarify that there is negative changes, negative because it is incompatible with professional ethics. You can refer to this in the book we have cited, which will be useful further.

A physician maximizes his or her utility with respect to y, t and s by choosing optimal amount. In that case maybe y star, t star and a star given the constant, constant will be as I already mentioned will be your induced supply function. h should be equalized with M upon delta plus s that is the total induced supply, h should be strictly positive, and t should have a minimum level of h and 1. So, the optima we can discuss one by one. This is based on the market condition and physician preferences.

A physician maximizes his/her utility $u(y, t, s)$ by choosing optimal amount (y^*, t^*, s^*) given constraints

$$h(\delta, s) = \frac{M}{\delta} + s, \qquad s \geq 0$$

and,

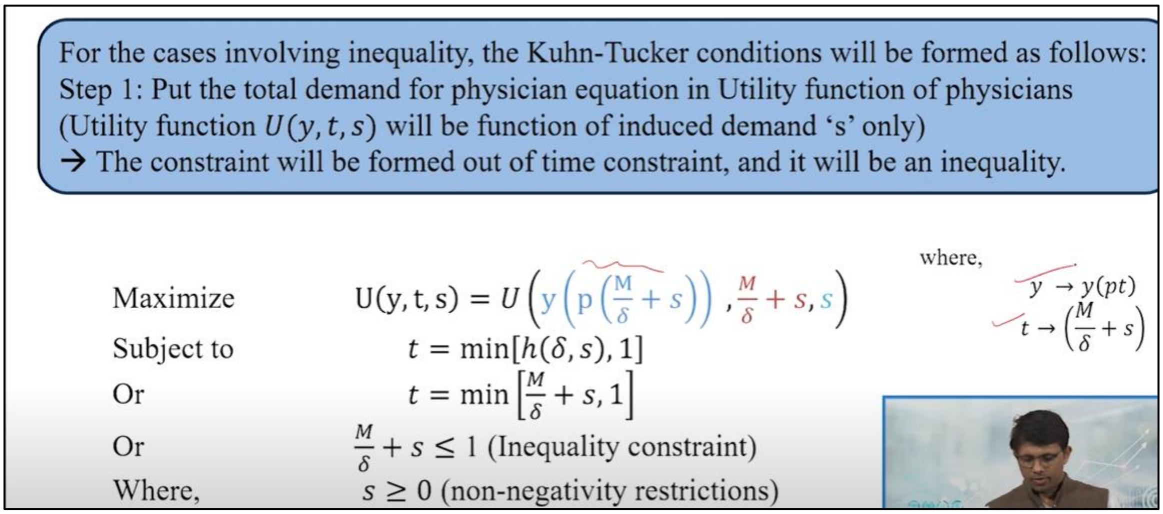
$$t = \min[h(\delta, s), 1]$$

So, all possibilities that mean s might be 0, s is induced supply, units might be 0, or t might be the maximum utilization or other boundary possibilities. s is greater than 0, t is less than 1, s is greater than 0, and t is equal to 1, etc. All four possibilities are discussed. So here, when the first boundary conditions occur, both are 1, meaning when physician density is low, market demand cannot be made with all physicians working the maximum amount of time. In the second case, inducement is 0, s equal to 0.

When ethics or preferences for laser are very strong and even physicians are working below capacity, they are only meeting primary demand. On the first case, t equal to 0 that means they are utilizing the maximum time, hence hardly there is any scope for the inducement. Then on the third case, that is called interior optima because both are the interior one, s is non-equal and also t is non-equal to 1. In that case, when physicians density is so high that even with optimum amount of demand inducement, they are working below full capacity. Last one is called optimum with t equal to 1 and s there is certain inducement, and physicians are inducing demand until full capacity is reached.

In the first one, these three require a simple Lagrangian optimization method, but there are some conditions that are inequal to the constraint function. Hence, non-linear programming is needed, and Kuhn Tucker condition is usually suggested. So these three cases require Kuhn Tucker conditions. The Kuhn Tucker conditions will be formed for the cases involving inequality. In the first step, we put the total demand for physicians' equations in utility functions that are U equal to y, t, and s, and this is a function of induced demand s only.

The constraint will be formed out of time constraints and will be an inequality. The utility function is y, t and s, y is y p t. So, y p t is explained over here. This is what is your t, and y is p t, and then this person is your t, then this is s and subject to p constraint is given, and inequality constraints are like that induced supply function should be at maximum utilizing one or less of the time and s is considered to be greater than equal to 0. Then only we are thinking of the Kuhn-Tucker conditions.



If we make that inequality conditions equalized with the boundary, that means if we add a certain dummy variable X, then that might be equalized with 1 that M upon delta plus s plus X we are just adding a dummy variable so that will be equalized with 1, where X is considered to be non-negative restrictions. So to consider a problem with non-negativity restrictions on the choice variables only with no other constraints. So, as an example, we are just maximizing profit or pi as a function of X subject to X being positive or greater than equal to non-negative. So, we are just taking non-negative restrictions on the choice variables. As a simple example, do you get optimized values in each case with a first-order derivative that is equal to 0? If not, then y you can just see.

If you just take the first order derivative since our constraint function is actually 0 and greater than 0 or non-negative. So in these functions you will find that the optimization occurs in first order derivative might occur at 0 place as well when X equal to 0. So this really dilutes the fact that even if X or the units of supply is 0 still held conditioning or the maximization point occurs at 0 output, which seems unreal and not right. And hence the optimization this way, simple optimization techniques is not clarifying the case. Hence, to explain all sorts of things, we will discuss these three cases: case 1, case 2, and case 3.

So, in case 1, the first-order derivative is 0. Where the output is the X is positive, whereas in case 2 and case 3 for case 2, the first order derivative is 0. Still, in case 3, the first order derivative is optimizing is tangent but not 0, it is less than 0. So when the first order derivative is 0, it gives value X equal to 0, which is really not the right indicator for optimizing. So, another solution to this is three cases, or in our typical way of optimizing conditions, through the first- and second-order derivatives, which we usually equalize to 0. But if you take the product in three cases, but product-wise, that is X times its first order value of the first order derivative, in all cases, this is optimizing and equal to 0. So, these can be considered as a solution for further use.

So, we will be utilizing when non-negative restrictions are imposed in equality constraint problem, the regular first order conditions becomes less than 0 or X is greater than 0 or the product of this t is equal to 0. The Lagrangian function will be formed by taking lambda as the Lagrangian multiplier. Subsequently, the first order conditions will be this: So, you have taken y, t and s with a Lagrangian constraint function and this is actually precisely M plus s, M plus s plus X equal to 1. So, we have taken 1 minus this, so this portion is actually equal to 0.

So, lambda times 0 is 0. So, the utility function is nothing but y, t and s. Through the Lagrangian optimizing formula, we used to take the conditions. In that case with respect to s, we know that like here with respect to s, with respect to X and with respect to lambda, three indicators are very important to discuss. So, when we take with respect to s, it is actually non-positive and whereas the product that is s time dZ by dS is equal to 0 and the product one is actually X times the faster derivative is equal to 0. We are going to use this, which we have started explaining here.

This second aspect that is the product is going to be very useful. We note that dZ by dX is nothing that is basically minus lambda. We use this information to simplify lambda once we define the value that is basically useful for simplifying the condition. Using this equation, we will get dZ upon ds is non-positive, and the product is 0. So, by solving this, we will find that I will solve these equations, usually called Kuhn-Tucker conditions.

We find $\frac{dZ}{d\lambda}$ and finally arrive into the solutions, which are the necessary conditions for the optimization problem. We will also clarify the exact interpretation of it. We will get n number of equations for n variable and to solve them to get optimal solution for the extreme order. And the first step is to confirm whether the extremum is maximum or minimum and we check the second order condition as given below. So, in that case when there are first order condition and second order condition, they all together decide the solution.

We used to take the bordered hessian determinant. We use a bordered hessian matrix represented as $\overline{H}$. A mixed matrix of first- and second-order derivatives of the objective and constraint equations is used to maximize the Z function and be subject to the constraint functions. You can follow the mathematical economics books to understand bordered Hessian determinants. Hence, in the four cases we saw that started explaining, they are starting with the boundary optimum, and then Kuhn-Tucker conditions will also re-emphasize this.

The solution would be with the simple example given that t is equal to 1; we have taken that t is equal to 1 in the first case. Hence, the y is equal to only yp, and t is no longer there; hence, the utility function is only y in terms of p. So, the effect of the increase in delta or number of services provided per person is q. So, the δ is very low, such that the $\frac{M}{\delta}$ should be greater than or equal to one. This means that primary demand exceeds the physician's working capacity with a constraint of t is less than or equal to 1 is binding. Since t is equal to 1 in our case and physician supply is measured in the physician's working time units, physician supply would be equal to **a**.

In this case, a times 1 unit per patient-physician supply equals a upon n, which is precisely the δ. Hence q is equal to a upon n, which is already mentioned as q in our example. Then, if and that is precisely the delta to take $\frac{dq}{d\delta} = 1$. This means that the billing of poor patients is proportional to physician density. So, once the physician density increases billing, the patients are considered proportional to it.

This is the first, second, third, and fourth conditions out of the Kuhn-Tucker formula, Kuhn Tucker conditions. We will find a number of things. We start with s equal to 0. Hence in a function, utility function you will see this. This is 0 is given rest as per our discussion we made in the previous one.

We want to say here that the effect of increase in the delta that is the doctors'-induced demand for health. The effect of this induced demand for health and the number of services provided per patient is that physicians do not induce demand because of a preference for leisure and ethical orientations. Hence s is equal to 0. Hence q is equal to M; whatever the q was there, this is not precisely the primary demand.

Hence dq by d delta is equal to 0. So, in this case billing's poor patients do not depend upon physician density in this delta range. In another interior optimum case, we will see that all have certain value. So, it isn't very easy. So, the optimum conditions will be p times first-order derivative and its y', u' and with respect to t and with respect to s that should be equal to 0.

So s is the function of exogenous variable. This case that is variables and s is a function of here delta p and M. So, in this case the demand is induced to the point where the marginal benefit of additional consumption equals the marginal utility loss of the additional working hours and the bad conscience resulting from demand inducement. The solution to it precisely you can see the same equation where just there is a minor change s plus delta time ds upon dq which we did in the previous two cases. So, there are two effects in this case.

One is direct and indirect effect. Indirect effect depends upon delta's impact on demand inducement at the level ds upon d delta. The last one again it has certain boundary conditions that t is used with the full capacity. So, in this case, s equals 1 minus M upon delta ($s = 1 - \frac{M}{\delta}$). So, physician density increases as delta increases or the induced demand increases. And effect of increase in this delta on the percent is that M upon; this is similar to that of the first case.

Here we have binding time constraint, but the delta is high. Hence, M upon delta is less than 1. This implies high motive for induced demand after the full capacity is due to high competition working below competition. Physician supply would be a times 1 units per percent, physician supply will be a upon n. And by all possibilities dq by d delta is equal to 1. That means billings per patient are proportional to physician density as long as all physicians work at full capacity.

So, you can see all three cases discussed in three regions. Until here we are discussing physician density and physician billing per percent that is Q. So, until the M level is reached, which is primary demand, you will see that physician billing per percent is proportionately higher. The first region of the straight line shows cases 1 and 4, in which the stage 1, stage 2, and stage 4 out of that 1 and 4 increase. In Region 2, the horizontal line shows case 2 without demand inducement exists, whereas case 3 in Region 3 has two direct and indirect effects.

So, that depends upon the indicators which we have already discussed. So, the theory of supply-induced demand that is SID is so far as SID is concerned, that is, SID is concerned with physicians' reaction to changes in physician density and the level of regulated fees that is P. It has an effect of physician density, which means physician density varies with cases of optimality and effect of fee level is ambiguous because substitution and income effects lead to a decrease in service value, respectively. There are some empirical identifications by different other authors. Fuchs, 1978 in his paper discussed a cross-sectional study based on USA between 1963 to 1970.

They observed that an increase of 10% in surgeons' density leads to a 3% increase in the number of surgeries, keeping other variables constant. This also supports the SID hypothesis we already discussed. Similarly, another case identified by some authors is on rising hysterectomy cases in India; some papers, some working papers can refer that some growth of these cases in private hospitals is higher than that of the growth of cases in public hospitals. This indicates supply induced units for health access or health demand and similar trends can also be observed in the C-sections, C- sections, deliveries in Indian context, especially highly cases are rising over the recent years and especially observed in private hospitals. SID is one explanation for increasing utilization of medical services per capita with increasing physician density, keeping prices constant. This indicates permanent excess demand, decreasing indirect cost, improve quality of treatment, defensive medicines, reverse causality, etc.

So, if permanent demand is higher, that d is much higher than the induced cases. That might be excess than that of the induced cases. Physicians have to turn away some patients. The left patients go to new physicians; this case corresponds to s equal to 0 and t equal to 1, which is the optimum. So, this basically affects supply extension at regulated prices and excess demand.

So, when excess demand exists, the supply curve tends to reach that level since demand is already there. That is why it is called permanent excess demand. It is largely affected by several reasons, maybe due to decreasing indirect cost, improve quality of treatment, new physicians in proximity means reduction in non-financial costs such as faster appointments, less travel time and time devoted to each case increases. Hence, demand shifts rightwards, and there is more new demand for a given amount of money or price. Different medicines are prescribed, such as medicine for the doctor's perspective that will work as defensive measures.

This is departing from best medical practices to minimize the likelihood of patient conflict, particularly in malpractice litigations. Reverse causality, if any, might induce the demand. New physicians are coming to places with more demand, and rather than demand, more physicians are coming. Consequently, high per capita medical service utilization leads to average physician density. Similarly, there are certain supply-side perspectives observed through racial discrimination by physicians. It is sometimes due to test-based discrimination, statistical discrimination, and efficient and inefficient discrimination.

So far as test-based consultations are concerned, many stereotypes are involved in our medical consultations. Doctors or the physician may or may not interact the patient because of their color or race. They may not treat subject to patients based on belief that they would not follow up medication or are not good candidate for treatment. So that is based on some statistical discrimination they make. Then, there are some discriminations based on welfare points of view, which are efficient and inefficient. One will shape up the physicians efficiently, and accordingly, more facilities are provided.

So, discrimination and health disparities are correlated. Other indicators, such as the wage rate and physician labor market, are essential. This helps in deciding the initial level or the decision for the physician to act as a physician or the decision to specialize and decide how much of work is there once a physician has a complete training. However, that training requires more time, and the demand in the market may not be cashed up to the supply because of the training period. Training of the physicians will show one chart that really time consuming and there are also entry barriers, eligibility criteria, etc., as part of entry barrier, then license to practice etc.

So, these also result in some monopoly rents that negatively affect the supply or the supply-induced demand. Another one might be positively impacted because of positive or better healthcare quality. So, this is the chart of the US, the time requirement in years of various residency and fellowship programs to act as a physician, starting with medical school, then different services, and the time age of a typical medical student. So, in different years, they will be capable of treating. We refer to Jay Bhattacharya et al. (2013), Health Economics, chapter 5. Similarly, physician wages, even if the training period of physicians is similar, their wages differ across countries, and supplies-induced demand is different due to salary differential.

Newly trained physicians may migrate to countries with better wages. Healthcare in India has become less accessible due to several cases of grants. So, this is what the countries are experiencing: doctor brain drain. The highest is noted in India, which has Indian cases because the wage rate is very different. In India, the UK, the US, Canada, and the UAE, the salary was in US dollars by PPP term in 2017, and India's figures are much lower.

Hence, doctors or physicians prefer to migrate to countries where they will earn better. We are referring to the Statista database, also available at our institute. You can also refer to the online documents for the record. I think these are all for physicians acting as a supply-induced healthcare demand.

In the following lecture, we will continue on the role of the physician as a health provider and discuss the role of hospitals as health providers. So, I think the references we have cited will be helpful. Thank you.