Paying on the Margin for Medical Care: Evidence from Breast Cancer Treatments

Reading Note 2, Yuzhe Wang

To solve the problem of increasing medical expenditures, a natural but not commonly accepted solution is the "top-up" insurance in which the consumers need to pay the additional cost of selecting more expensive treatments relative to the cost of baseline treatments. Besides "top-up" insurance, there are two common types of insurance called "full coverage" and the "no top-up". The "full coverage" insurance is a design in which consumers do not need to pay the incremental costs of more expensive treatments. The "no top-up" insurance is a design in which only "cost-effective" treatments are covered. Compared to the "full coverage" and the "no top-up" insurance, the "top up" insurance is regarded as a compromise that is more efficient in categorizing patients over treatments by internalizing the marginal treatment costs. In this paper, the authors firstly introduced the framework of welfare consequences of the three types of insurance by demonstrating the demand curve for the more expensive treatment, and then estimated the demand curve to quantify the welfare effects of the "top-up" insurance by using the logit model.

The article used the data on breast cancer treatments, which was perfect in estimating the demand curve because of two types of treatment: mastectomy (M) and lumpectomy (L). M was cheaper and there was no survival rate difference between M and L. In principle, the cost of M would be fully covered by insurance while L would not. Therefore, the data can firstly help to identify the demand curve of L by estimating the relationship between the cost variation between L and M (relative price of L) and the share of L, and then the authors calculated the empirical welfare effects of "top-up" insurance in this breast cancer case.

To estimate the relative price of L, the authors firstly estimated the patient's relative utility of L. To estimate the relative utility of L, the authors firstly explored the factors which could influence the choice of treatment. In other medical literature, there was an empirical link between distance (travel time) from radiation facilities (patient need to take radiation therapy in L) and the treatment choice (M or L). The initial evidence from the data also showed that the longer the travel time was, the less patients were likely to choose L and the relation was not sensitive to the inclusion of other control variables. After identifying the significant relationship between the distance and the choice of treatment, the authors built logit models to estimate the relative utility of L.

In the benchmark logit model, the regressor was the relative utility of L and the regressand was the distance. When the relative utility of L was larger than 0, the patient chose L. Besides the benchmark logit model, some logit models also included other control variables and the interaction between the control variables and the distance. Apart from these standard logit models, the authors also included the random-coefficient logit model by assuming that coefficients on distance followed the lognormal distribution. The empirical results of the standard logit models showed that when the travel time was 10 minutes longer the patient was less likely to choose L by around 0.7 to 1.1 percent. The negative relation was stronger in the random-coefficient model but the estimation was very noisy. In conclusion, the results of the logit models showed a robust negative relationship between the relative utility of L and the distance (travel time to radiation facility).

After estimating the relative utility of L, the authors used the estimation to define the relative price of L (also called willingness to pay for L in the article) and obtained the demand curve of L. In this way, each patient in the dataset would have an estimated relative value of L and the distribution of the relative price of L was obtained. When taking the inverse function of the distribution, the article gained the demand curve of L by showing the relationship between the relative price of L and the share of the choice of L.

After estimating the demand curve of L, the welfare effects of "top-up" insurance were calculated. To compare welfare, the authors firstly defined the three types of insurance on the demand curve. When the relative price of L was 0, the patient was regarded to take the "full coverage" insurance because there was no difference between M and L. When the relative price was equal to the incremental payment of L, the "top-up" was denoted because the patient internalized the marginal cost of L. When the relative price was the total cost of L, it was the "no top-up" because the treatment L was not covered at all. In principle, compared to the "top-up", the "full coverage" leads to welfare loss because the patients who cannot pay the incremental fee still choose L (paid by the insurer), and the "no top-up" also brings welfare loss because patients who prefer L will choose M to avoid the total cost of L. The empirical results also showed that, for example, when the incremental payment is \$10,000, the "top-up" decreased the rate of L by 35%-40% and increased the total welfare from \$700 to \$2,500 per patient compared to "full coverage", and increased the rate of L by 15%-25% and increased the total welfare from \$700 to \$1,800 per patient compared to "no top-up".

The previous analysis only considers the ex-post efficiency and the authors further discussed the ex-ante welfare consequences of the three types of insurance. Similar to the previous methods, the author firstly defined the ex-ante utility which was given by the solution when maximizing the expected utility with the CARA utility function. The results showed that, in each level of risk aversion, compared to "full coverage", both "top-up" and "no top-up" reduced the insurer's cost and the consumer surplus, and "no top-up" also increased the social cost relative to full coverage. The more interesting result was that the social cost of "top-up" relative to "full coverage" changed with the level of risk aversion. For the median level of risk aversion, "top-up" increased social welfare, but for a higher level of risk aversion, it decreased the social welfare compared to "full coverage". Because the "top-up" insurance did not win the "full coverage" insurance in all levels of risk aversion, the authors provided two additional insurance contracts. The first one was called "first best" in which the insurer covered costs of M and pad a lump sum of indemnity which was continuous between 0 and the incremental cost of L. Although "first best" insurance reduced the social cost relative to "full coverage", it was impractical due to the continuum of indemnity and the form of lump sum payment. The author then provided a more "practical" design in which the insurer only covered part of the "top-up" coverage. Although this design could meet the problem of adverse selection, it was less significant in the data used in the paper and the results also showed a reduction in social cost compared to "full coverage".

In conclusion, the paper mainly demonstrated the welfare gain from the "top-up" insurance by estimating the demand curve of the more expensive choice of treatments. In addition, the paper also discussed the ex-ante efficiency of the "top-up" insurance and illustrated that in the simple exercise the "top-up" insurance might not dominate the "full coverage" insurance in a higher level of risk aversion. More importantly, though the paper only considered the survival benefits of the patients and did not consider other factors such as "body integrity", it still provided a framework to analyze the welfare consequences of "top-up" insurance and contributed to the empirical strategy to monetize the additional cost of the more expensive treatment option.