

Reading Note: Einav, Finkelstein, and Williams – *Paying on the Margin for Medical Care* (2016)

In many high-income countries, two primary insurance policies are available to the public. In the United States, there is a “full-coverage” insurance design where patients generally do not pay the incremental cost of choosing a more expensive treatment. Alternatively, in countries like the United Kingdom, a “no top-up” insurance design is employed where predetermined “cost-effective” treatments are fully covered but patients must pay for all other treatments. In both insurance designs, patients inefficiently sort into different treatments, leading to substantial social welfare costs. The authors of *Paying on the Margin for Medical Care* proposed an alternative “top-up” insurance design where patients pay the incremental cost of the more expensive treatment and argued that, in most cases, it would lead to more efficient sorting into treatments as well as welfare gains over the traditional insurance designs.

This paper took a unique approach to address this problem by using breast cancer treatments as its object of interest. Breast cancer has two primary treatment options: mastectomies that remove the entire cancerous breast and lumpectomies that remove the tumor while preserving the breast. Not only are the treatments different, but lumpectomies are also followed up by a course of 25 visits to a radiation facility while mastectomies do not have similar post-surgery followups. While the costs of both treatments are similar, the authors assumed that the recurring visits to a radiation facility can be monetized, making the total cost of a lumpectomy much higher than that of a mastectomy. This assumption allowed the authors to abstract away from issues regarding monetizing adjusted years of life between these two treatments, making the analysis more straightforward as they implied that both treatments had similar levels of quality¹ but varying price levels. In their proposed “top-up” design, the baseline treatment, the mastectomy, is fully covered, but the patient must pay the incremental cost of the lumpectomy. By their assumptions, the monetization of travel for radiation therapy was analogous to the incremental costs patients would have to pay for a lumpectomy in their proposed insurance design. The authors argued that by having the patients internalize the marginal treatment costs, there would be a more efficient sorting of patients across the two treatments based on their relative valuations. In their quantitative analyses, the authors used data from the California Cancer Registry² which included treatment decisions of breast cancer patients in California, alongside demographic and diagnostic information, between 1997 and 2009; they supplemented this data with the distance to the nearest radiation facility for each patient.

Before doing welfare analyses, the authors examined the relationship between treatment decisions and the distance to the nearest radiation facility (as it was analogous to incremental costs). The authors used a logit regression model, with various specifications, to investigate how distance affected the likelihood of choosing a specific treatment. By calculating the average marginal effects with bootstrapped standard errors, they saw that the effect of distance was statistically significant and stable across various specifications. Overall, they found that increasing the time to the nearest radiation facility by ten minutes reduced the likelihood of a patient choosing a lumpectomy by roughly 0.7 to 1.1 percentage points. Afterward, the authors estimated the demand that illustrated the relative willingness to pay for lumpectomies compared to the baseline mastectomy. This estimation was done by comparing each patient’s valuation of the two treatment options (v_i ³) and using the cumulative density of these valuations to estimate the demand curve to see what proportion of treatments are lumpectomies at various price levels. With their proposed “top-up” insurance design, patients who valued the lumpectomy over its marginal cost would get the lumpectomy whereas patients who did not value the treatment as high would get the mastectomy instead.

¹ The quality of treatments was determined to be similar as the survival outcomes did not vary significantly between treatments.

² The medical information for each patient is directly collected from the patient’s medical records when they are diagnosed with cancer.

³ v_i is positive if the patient values the lumpectomy more than the mastectomy.

After the authors constructed a demand curve for lumpectomies, they conducted ex-post welfare analyses for the three insurance designs where they approximated the incremental cost of a lumpectomy to be \$10,000⁴. The authors noted that their proposed design created more efficient sorting of patients into the two treatments – their proposed design increased the lumpectomy rate by 15-25 percentage points relative to the “no top-up” design and decreased the lumpectomy rate by 35-40 percentage points relative to the “full coverage” design. Furthermore, the authors saw that the “full coverage” design not only produced inefficient treatment decisions (as patients who have lower values of v_i inefficiently chose lumpectomies) but also had welfare costs of at least \$710 per patient relative to the efficient allocation; similarly, the authors saw that a “no top-up” design also had welfare costs of at least \$800⁵ relative to the efficient allocation and led to inefficient sorting as well. From an ex-post perspective, it seemed as though their proposed design yielded better outcomes than the two traditional insurance designs.

However, the authors noted that ex-post welfare analyses were insufficient as they may leave out “a part of the picture” so they also briefly looked at ex-ante efficiency as well. Using a standard CARA utility function, the authors examined ex-ante risk exposure (π_i) and saw that the “full coverage” insurance design’s risk exposure was zero. They also noted that, while the ex-ante risk exposure for their proposed insurance design was positive, it was still less than that of the “no top-up” design. After looking at ex-ante efficiency, the authors concluded that the “full coverage” design was better than the “no top-up” design at all levels of risk aversion. However, the “full coverage” design’s social welfare was higher than that of their proposed design for high enough levels of risk aversion. Hence, the ex-ante relative ranking between designs was somewhat dependent on the level of risk aversion, which meant that determining the “best” design was inconclusive when looking at ex-ante efficiency alone. However, it is important to note that if the authors did not look at ex-ante efficiency, they would have missed the risk exposure generated by their proposed design. This generated risk exposure could be more costly than its allocative proficiency, highlighting that it is not a “perfect” solution to the problem but still better than the current insurance designs.

The authors concluded that their proposed “top-up” design was optimal because it created efficient sorting into treatments when looking at ex-post efficiency. Furthermore, when looking at ex-ante efficiency, their proposed design also maximized social welfare when compared with a “no top-up” design and, in some cases, a “full coverage” design as well. This proposed design is a middle ground that efficiently sorts patients by not overpricing more expensive treatments. While this paper provided a robust framework for analyzing insurance designs, there is one issue to address. The paper relied on the relationship between treatment decisions and incremental costs (via the distance to the nearest treatment facility) but one problem arises: they don’t know which treatment facility lumpectomy patients received treatment from – the closest facility may not be their actual radiation treatment facility. Suppose the authors could find which facilities the lumpectomy patients received radiation treatment from (or if patients moved to be closer to a particular facility); in that case, it may alter their findings as there can be a dimension of quality (or other factors) that could impact treatment decisions alongside incremental costs (through distance to the facility). Potentially incorporating fixed effects to capture facility quality could strengthen the authors’ methodology and potentially find a more accurate relationship between price and treatment decisions that net out other factors, aside from costs, that also affect treatment decisions.

⁴ The “total” cost of the lumpectomy was approximated to be \$50,000 as the baseline mastectomy cost was roughly \$40,000.

⁵ These associated welfare costs are based on the “richest” model which allows for heterogeneity in price responses. In the simpler models, the costs tended to rise up to \$2,000 and \$1,400, respectively.