

1 Data

We now turn to the empirical section of this paper. To estimate our models, we will use administrative data provided by FONASA from all sick leaves issued from January 2018 to October 2022 in the public health system. We consider authorized sick leaves only. Our sample includes 48,611 different physicians and 1,445,696 different patients. The dataset contains limited information: each observations consists of a single issued sick leave including encrypted ID’s for the physician and patient involved (as well as the patient’s employer), duration of sick leave and *cie10* classification of the condition, and some data on issuance and payment methods.

Table 1 presents descriptive statistics on sick leaves by physician and patients, i.e. on the empirical distributions of X_j and X_i . One first indication of the skewness of both distributions is the difference between mean and median: whereas the median physician issued 24 sick leaves over the sampled period, the *average* physician issued 88. The right tail is long, the 90th percentile physician issued 216 sick leaves, which quickly goes up to 1,007 at the 99th percentile, and then the most issued by a single physician is more than eight times that amount at 8,743.

The story is similar with regard to patients, as a considerable share of the sample appear only once and then the median is just *two* sick leaves received over the sampled period, goes up to 7 and 14 at the 90th and 99th percentiles, respectively, but then the single worker with the most sick leaves authorized is over five times the 99th percentile at 73.

Statistic	Physicians	Patients
Mean	88	3
S.D.	199	3
Min	1	1
Max	8,743	73
Percentiles		
10th	2	1
50th	24	2
90th	216	7
99th	1,007	14

Table 1: Total sick leaves issued *by* physicians *to* patients
Summary statistics over the sampled period

Long tails are not only present in both distributions, but interrelated. ? argues that there is a meaningful degree of assortative matching in this sample, where physicians willing to grant many sick leaves are matched with—found by—patients with a high demand for them.

This effect is captured by our framework: patients with serious conditions prioritize

higher “quality” physicians, who, in turn, tend to treat patients with greater medical need (κ_i) due to their endogenously determined higher strictness threshold ($\bar{\kappa}_j$). Conversely, patients with lower κ_i , but a strong desire for sick leave (γ_i), naturally gravitate toward more lenient physicians with lower strictness thresholds ($\bar{\kappa}_j$).

We don’t put the *cie10* classification to use to estimate the distribution of κ , partly because κ_i ’s interpretation as willingness-to-pay isn’t one to one with strict medical severity, but mainly because the very subject of this paper is fraud, and in that regard the health status of patients reported by physicians are unreliable, especially for the most common conditions attested to by physicians (see Table ?? in the Appendix ??). In fact, these routine conditions are in many cases hard to gage even for well-intentioned physicians. ? establish that physicians in these scenarios have no incentive to place doubt on the patients’ self-reported symptoms. This is part of the reason we choose to fit a structural model rather than opting for more straightforward, data-oriented methods of estimation.¹

¹As the severity of the ailment increases, and the duration of sick leave accordingly, higher institutional and medical scrutiny is placed on diagnosis and issuance. We will discuss this in the next section.