Understanding FTE Data - Texas

Part I: General to All States

FTE (Full-time equivalent), is a unit that indicates the workload of an employed person in a way that makes workloads or class loads comparable across various contexts. An FTE of 1.0 is equivalent to a full-time worker, while an FTE of 0.5 signals half of a full work load. What is in the FTE files of our project, however, is a conditional use of this concept.

What do the head values mean -- There is no primary data regarding FTE. For simplicity, dentists are counted as one head. If they appear in multiple locations, the FTE for each location is then one divided by the number of appearances. However, when I counted the heads and assigned their head values, the Medicaid and CHIP files are separate. The providers who appear only in Medicaid or CHIP will have only one head, while those appear in both will have two heads. Due to the lack of a name column in the final output, such details are lost. Since the hygienists' head values are derived from dentists, they are also affected accordingly.

Example: the screen shot below is of the example (from Colorado) of Danny Abboud. Danny appears in 4 addresses in Medicaid and he has .25 for each. Then he appears in CHIP and his head was counted as one a second time and he got .33 for each of the three locations (these three locations overlap with Medicaid).

Medicaid listing

ABBOUD, DANNY	13065 E 17TH AVE STE F840,AURORA,CO,80045-2533	0.250000
	1610 CANYON BLVD,BOULDER,CO,80302-5407	0.250000
	3216 ARAPAHOE AVE,BOULDER,CO,80303-1043	0.250000
	7735 W LONG DR UNIT 9,LITTLETON,CO,80123-1262	0.250000

CHIP listing

ABBOUD, DANNY	1610 CANYON BLVD,BOULDER,CO,80302-5407	0.333333
	3216 ARAPAHOE AVE,BOULDER,CO,80303-1043	0.333333
	7735 W LONG DR UNIT 9,LITTLETON,CO,80123-1262	0.333333

In the all-inclusive Medicaid and CHIP combined file (the one produced by the outer join), it is possible to derive the Medicaid and CHIP combined value of a single location by summing the two relevant columns. However, it is very important to take into account the fact that each dentist who accepts both has been counted twice and thus has two heads and these two heads may each have different locations or may not have equal number of locations between Medicaid and CHIP.

Furthermore, the one head count in effect equals the value of the portion of time a dentist might spend seeing Medicaid and/or CHIP patients (which is unknown and cannot be derived from IKN data), assuming each dentist also spends time (or spends the larger portion of his/her time) seeing other patients who have private insurance.

Uncertainty of which location(s) a provider might practice at – IKN lists each dentist at all locations of his/her practice without specifying which locations a dentist would actually see patients.

For example, Dr. COLBERT, MILDRED CHARISSE, has multiple rows in the Medicaid FTE file, including locations of Bridge City, Beaumont, Galveston, Port Arthur (all in Texas). According to the website Gulfside Dental Orthodontics (https://gulfsidedental.com/about/meet-the-team/), she only sees patients at the Beaumont location. Without such details from IKN data, the FTE files take in the IKN listings at face value and counts Dr. Colbert as seeing patients at all locations owned by Gulfside Dental, splitting her head value across the locations.

Data Irregularities

Differences in the strings of addresses caused by the different ways an address is abbreviated -- In the FTE files, each address is considered unique in the current FTE file as Pandas looks at string patterns, not the meaning of the addresses.

Some addresses have variations, such as Dr./Drive. Some have the practice name ahead of the street address. This causes the head values of providers in the same practice to be each listed in different/separate rows, which is a data feature that can only be corrected manually.

Example -- These two addresses have slight variations (highlighted part) but refer to the same location. When the names are retained, you can see that all three rows are distinct.

15901 E BRIARWOOD CIR UNIT 350, AURORA,CO,80016	general	MINGLE,MICHAEL	1
15901 E BRIARWOOD CIR UNIT 350, AURORA,CO,80016	general	RANKIN,JENNIFER	1
15901 E. BRIARWOOD CIRCLE SUITE 350, AURORA,CO,80016	general	YORK-JOHNSON,KIMBERLY	1

However, in the final version, you will see two rows like this
15901 E BRIARWOOD CIR UNIT 350, AURORA,CO,80016 general 2
15901 E. BRIARWOOD CIRCLE SUITE 350, AURORA,CO,80016 general 1

A similar problem can occur with GPS values. Here is an example from NJ in Tina's

1 BRITTON PL STE 11, VOORHEES, NJ, 08043 39.8451463,-75.00344729999998
1 BRITTON PLACE, VOORHEES, NJ, 08043 39.8447495,-75.00326779999997

This address refers to the same location, and there is not only a variation in the address strings, but also in the GPS strings.

If an address has exact IKN address-GPS combinations in multiple rows, they are true duplicates. If they occurred in the FTE final output, it is due to the existence of an extra space in either the X or the Y columns (not visually noticeable), causing pandas to interpret each as unique. I dropped off duplicates using only IKN_addr as the subset (in the case that GPS is used as index when reshaping columns, a different routine of dropping off duplicates is needed).

Hygienists

The column "hygienists" (if present) contains values for hygienists who practice independently and thus have submitted claims. Very few are matched and thus most (but not all) cells have 0s.

The other two columns, "hygienists for general" and "hygienists for pediatric" contain derived values, each by multiplying the value of the dentist's column by 2, essentially assuming that each dentist has two hygienists on average (per Dr. Serban's previous practice).

Head counts in the state report vs. FTE

The head value or sum of head values in the FTE file is not necessarily going to match the count of heads in the State Report tables. The statistical table numbers simply say that, 'in the provider full name column, there are x number of dentists' names', while an FTE row is the sum of fractional values derived from dividing a dentist's head by the number of his appearances (how many IKN rows). A further complication is that the heads are counted twice for those who accepts both Medicaid and CHIP.

These are two sets of numerical data that are derived differently even though they should in theory match.

Part II: Specific to Texas

BOD

BOD data was downloaded Jan 8, 2020 from

http://tsbde.texas.gov/resources/licensee-lists/

BOD data is relatively good, contains county column which the state report can use.

IKN and NPPES

7.4 % of Medicaid rows do not have names but have NPIs; of these 7% could not be matched with NPPES.

33.5% of CHIP rows do not have names, of these, 0.7% of CHIP rows were not matched. These rows that have numbers in the IKN NAT_PROV_IDENT column but cannot be matched with NPPES were put into the pool of a direct match between IKN and BOD (some of them were eventually matched this way).

Of the rows that do not have dentists' names but have NPIs in the IKN NAT_PROV_IDENT column and were matched with NPPES, the group practice names were copied over to the provider full name column and used as provider names. A sample check with NPPES names reveals that these IKN practice names often are the same as the NPPES names that have been matched.

1936 out of 95380 rows (2%), or 210 out of 11345 of the unique IKN names (1.8%), contain dentists who has a non-TX-BOD address but a TX address in NPPES. These were retained so as to be consistent with Colorado and Florida.