The following decisions were made in organizing the data:

1. We discarded the additional information in our data which we did not use in our schema. This includes, for instance, the website of each public health unit, and the date on which patients' test results came back.

2. We deleted items with missing pieces of information, which would have resulted in null values, violating our schema(*outcome* table, *patient* table).

3. Similarly, we deleted rows with entries like 'UNKNOWN' and 'NO INFO/MISSING', to conform the data to our schema((*exposure* table).

4. Some values in the *property* file appeared more than once. In such cases, the maximum price for each real estate property was chosen as the nominal price and the duplicates were eliminated.

5. Where the *property* csv file made finer gradations in area names, such as 'Downtown Toronto', 'Midtown Toronto', and 'Toronto', we simplified these to a coarser form, in this case 'Toronto', so as to conform to the convention used by the *phu* csv file.

6. We deleted some data in *phu* csv file, where its city is not shown in *property* csv file, which violet the referential integrity constraint.