Some notes from meeting with Tanima and Paul:

1. ~90-95% of all NPI's and prov\_unique should be in one-to-one correspondence

2. Possible reasons for multiple-to-one NPI and prov\_unique relationship:

(one NPI has multiple prov\_unique's)

1) A prov\_unique may refer to an individual, a group, or a facility, etc. So an NPI may correspond to a prov\_unique for group practice, and a prov\_unique for an individual.

2) One NPI may correspond to a prov\_unique for a supervisor and a prov\_unique for a supervisee. Examples of a supervisee are a nurse practitioner, or a physician assistant. A nurse practitioner may prescribe drugs. A physician assistant may prescribe drugs in some states.

(Others)

3) Some NPI's may be recorded by error

Possible ways to solve the problem:

1) Depending on type of the prescribed drug, we decide whether to use prov\_unique to obtain provider level information.

2) Treat the one-NPI-to-multiple-prov\_unique’s as unknown.

3. Some trivial NPI’s to be treated as missing/unknown:

1) “5”

2) “L”

3) “LLLLLLLLLLL”

4) “G”

5) “” (empty)

Actually only “L”, “LLLLLLLLLLL”, and “” are found in the data. The other two are not found.

4. Provcat was created by OptumInsight to label providers. It may be unreliable information.

5. Taxonomy 1 and Taxonomy 2 are provider-reported primary specialty and secondary specialty respectively, and may be more reliable than provcat.

6. If two prescriptions have the same NPI but different prov\_unique's, and further if the two prescriptions have the same claimID, then the two prescriptions are likely to go under the same provider.

7. The standard cost from pharm data sets is a standardized charge amount, adjusting for drug types, states, copays, etc using an unrevealed algorithm. Higher standard cost implies higher charges, but the actual value does not mean the actual charge.

8. Connections between medical data and confinement data:

1) Medical data sets contain both inpatient and outpatient claims.

2) Inpatient claims from medical data sets are summarized into claims in confinement data sets. The unsummarized and summarized claims are linked through a unique confinementID. This implies that summarized inpatient records from medical data are a subset of confinement data.

3) Some claims are found in confinement data but not in medical data. Rare diseases may be examples. A possible reason is that inpatient claims went into confinement data and were submitted to the insurance company. The claims got denied by the insurance and thus did not enter medical data. However the claims might have stayed in confinement data for so long that people forgot to remove them from confinement data.