CLAIMS

What is claimed is:

1. A method comprising:

analyzing, by a computing system including processing circuitry and memory, one or more data tables stored by a data repository with respect to one or more criteria, the one or more data tables including insurance claims data corresponding to one or more treatments provided to an individual for a biological condition and the one or more criteria indicating one or more formats of insurance code identifiers;

determining, by the computing system, a plurality of insurance code identifiers included in the one or more data tables that corresponds to the one or more criteria,

analyzing, by the computing system, a subset of the plurality of insurance code identifiers to determine one or more formats of the subset of the plurality of insurance code identifiers;

generating, by the computing system, one or more requests of an application programming interface (API) calls that include an insurance code identifier of the subset of the plurality of insurance code identifiers;

obtaining, by the computing system and in response to the one or more requests of the API, a data file that includes information corresponding to the insurance code identifier;

extracting, by the computing system, a treatment identifier from the data file; and

generating, by the computing system, an additional data table with a row indicating that the insurance code identifier corresponds to the treatment identifier.

2. The method of claim 1, comprising determining, by the computing system, one or more columns of the one or more data tables that include insurance code identifiers corresponding to treatment of individuals.

3. The method of claim 1 or 2, comprising:

determining, by the computing system, that a first insurance code identifier of the plurality of insurance code identifiers is a duplicate of a second insurance code identifier of the plurality of insurance code identifiers; and

removing, by the computing system, the second insurance code identifier from the plurality of insurance code identifiers to produce the plurality of insurance code identifiers.

4. The method of any one of claims 1-3, wherein the plurality of insurance code identifiers correspond to a plurality of National Drug Code (NDC) identifiers.

5. The method of any one of claims 1-4, comprising:

identifying, by the computing system, a first insurance code identifier of the subset of the plurality of insurance code identifiers;

determining, by the computing system, that the first insurance code identifier corresponds to a first format of insurance code identifiers; and

modifying, by the computing system, the first insurance code identifier to produce a second insurance code identifier that corresponds to a second format of insurance code identifiers.

6. The method of claim 5, comprising:

generating, by the computing system, one or more additional requests of an additional API that include the second insurance code identifier;

obtaining, by the computing system and in response to the one or more additional calls of the additional API, an additional data file that includes information corresponding to the third insurance code identifier; and

extracting, by the computing system, the insurance code identifier from the additional data file.

7. The method of any one of claims 1-4, comprising:

generating, by the computing system, one or more additional calls of the API that include an additional insurance code identifier;

obtaining, by the computing system and in response to the one or more additional calls of the API, an additional data file that includes additional information corresponding to the additional insurance code identifier; and

determining, by the computing system, that at least one valid insurance code identifier is

not included in the additional information.

8. The method of any one of claims 1-7, comprising:

generating, by the computing system, one or more additional calls of the API that include an additional insurance code identifier of the subset of the plurality of insurance code identifiers;

obtaining, by the computing system and in response to the one or more additional calls of the API, an additional data file that includes additional information corresponding to the additional insurance code identifier;

analyzing, by the computing system, the additional information with respect to one or more additional criteria; and

determining, by the computing system, that the additional information does not include at least one drug identifier.

9. The method of any one of claims 1-8, comprising:

generating, by the computing system, one or more additional calls of an additional API that includes the treatment identifier;

obtaining, by the computing system and in response to the one or more additional calls, an additional data file that includes additional information corresponding to the drug identifier; and

determining, by the computing system and based on the additional information, a class of drugs that corresponds to the drug identifier.

10. The method of any one of claims 1-9, comprising:

determining, by the computing system and based on the additional information, a source of the class of drugs

analyzing, by the computing system, the source of the class of drugs with respect to a prioritized number of drug sources;

extracting, by the computing system and from one or more fields of the data file, at least a portion of the additional information; and

adding, by the computing system, the at least a portion of the additional information to the row of the database table.

11. The method of claim 10, wherein the prioritized number of drug sources includes a first source of the class of drugs having a first priority and a second source of the class of drugs having a second priority that is lower than the first priority.

12. The method of claim 10 or 11, comprising:

analyzing, by the computing system, the additional information with respect to the first source of the class of drugs;

determining, by the computing system, that the additional information does not include a source of the class of drugs that corresponds to the first class;

analyzing, by the computing system, the additional information with respect to the second sources of the class of drugs; and

extracting, by the computing system, at least a portion of the additional information from the additional data file in relation to the second source of the class of drugs.

13. The method of any one of claims 1-12, comprising:

receiving, by the computing system, a request to identify a group of individuals that received a treatment in response to a biological condition being present with respect to the group of individuals, wherein the request includes a name of the treatment or a class of the treatment;

analyzing, by the computing system, one or more values of one or more rows of the additional data table to determine one or more insurance code identifiers that correspond to the name of the treatment or the class of the treatment;

analyzing, by the computing system, values of a number of rows of the one or more data tables to determine one or more rows that include the one or more insurance code identifiers; and

determining, by the computing system, one or more identifiers of individuals included in the one or more rows to produce a cohort of individuals that received the treatment in relation to the biological condition.

14. The method of claim 13, comprising:

determining, by the computing system, genomics information of the cohort of individuals

that received the treatment in relation to the biological condition; and

applying, by the computing system, at least one of one or more statistical techniques or one or more machine learning techniques to determine one or more features of the cohort of individuals.

15. The method of claim 14, wherein the one or more features include at least one of a genetic mutation included in respective genomes of individuals included in the cohort of individuals, a genetic mutation of cell-free deoxyribonucleic acids (DNA) included in one or more samples obtained from individuals included in the cohort of individuals, an amount of cell-free DNA having the genetic mutation for the respective individuals included in the cohort of individuals, or a change in the amount of cell-free DNA having the genetic mutation for the respective individuals included in the cohort of individuals over a period of time.

16. A system comprising:

one or more hardware processing units;

one or more computer-readable storage media storing computer-executable instructions that, when executed by the one or more hardware processing units, cause the system to perform the steps of the method of any one of claims 1 to 15.

17. One or more non-transitory computer-readable storage media storing computer-executable instructions that, when executed by one or more hardware processing units, cause a system to perform the steps of the method of any one of claims 1 to 15.