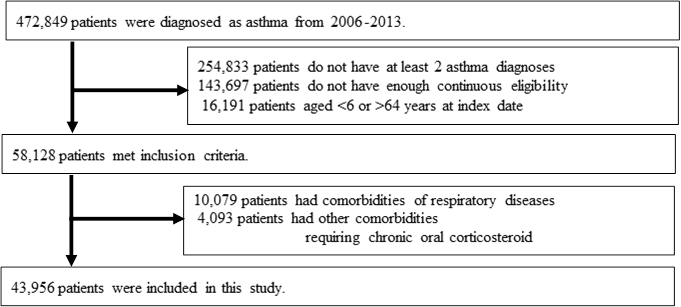
**Data checking for asthma treatment step project**

**Questions to Richard**

1. Patient flow chart: request updated patient flow-chart (after considering below concerns).

This is an example of the flow we need.



1. Question on study period requirements. We requested data from 2006-2013 but some date variables are out of the eligible study period.
   1. asthma\_index\_date: is from 01jan2006 through 29jun2012. However, we require at least 6 months pre-period eligibility. Thus, the index date should start no earlier than 01july2006. From prior analysis request document, the following may be helpful, “(NOTE: asthma diagnosis can occur prior to the index date, however, these diagnoses would not have 6 months eligibility prior to the dates.)” Recall that we want to include both incident and prevalent asthma patients.
   2. Elig\_days: minimum value is 28 days. However, we requested that each participant have at least 2.5 years of continuous eligibility.
   3. no\_cont\_eligi\_fu: why do we have negative values in this variable? We were not expecting them.
   4. no\_dx: we requested a number of variables within the baseline characteristics be measured over the 6-month pre-period. We are concerned that some did not have this length of pre-period (perhaps less than 6 months) and therefore, such variables like no\_dx will not make sense.
      1. See other red variables within the baseline characteristics table for more concerns like that mentioned for no\_dx.
   5. no\_med\_visit: concerned that max is 179 for number of medical visits. Was medical visit defined as a separate medical encounter?
   6. no\_procedure: also seems high, but could be appropriate – could you double check this?
2. We thought that the number of participants would be consistent across the files, however, N in the exposure file is less than others.
   1. Within the Exposure Dataset, elig\_stop variable is not consistent with the same variable in the baseline characteristics file. Can we please have the same N in each of the three databases? Hopefully, this will solve the elig\_stop inconsistency.
3. Other Exposure database variable concerns:
   1. step\_start: concerned that the max date is beyond that of the study period.
   2. step\_stop: concerned that the max date is beyond that of the study period.
4. Outcomes look clinically reasonable – thanks!
   1. Assuming that if the cohort N changes, then all databases including outcomes will be updated based on the new pre- and post-periods (i.e. eventdt: min of 01jan2006 is not possible as this would be in the pre-period).
5. **Baseline Characteristics**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Format** | **Assessment**  **(Correct? Y/N)** | **Note** |
| pat\_id | 0268AAAAAAAACPBG | Y | N = 60,479 |
| der\_sex | F/M/U | Y | There is only two observations with U,  58.2%F and 41.8%M |
| der\_yob | YYYY | Y | Min:1942, Max:2006 |
| pat\_region | E/MW/S/W | Y | E:28.76%, MW:31.32%, S: 27.12%, and W: 12.80% |
| index\_asthma\_date | 01jan2006 | N | **Min**: 01jan2006, **Max**: 29jun2012  Why are there index date before 01Jul2006 because we need at least 6 months eligibility before the index date? ***(Inclusion criteria#2)*** |
| fst\_asthma\_dx | 01jan2006 | N | Same as index\_asthma\_date |
| second\_asthma\_dx | 02jan2006 | Y | **Min**: 02jan2006, **Max**: 31may2013 |
| fst\_asthma\_dx\_year | 2006 | Y | **Year** 2006 – 2012 with reasonable distribution |
| second\_asthma\_dx\_year | 2006 | Y | **Year** 2006 – 2013 with reasonable distribution |
| age | 6 | Y | **Mean (sd):** 31.24 (18.21), **Min:**6, **Max:**69 |
| region | 1/2/3/4 | Y | Same as pat\_region |
| season\_index | 1/2/3/4 | Y | [1 = summer (21.82%), 2 = fall (25.77%), 3 = winter (26.78%), 4 = spring (25.62%)] |
| gender | 0/1/2 | Y | Same as der\_sex |
| elig\_start | 02jan2006 | Y | **Min**: 01jan2001, **Max**: 01dec2011 |
| elig\_stop | 02jan2006 | Y | **Min**: 31jan2006, **Max**: 30jun2014 |
| elig\_days | 28 | N | **Mean (sd):** 2340 (870), **Min:**28, **Max:**4929  The eligibility should be at least 2.5 years (910 days) as our criteria requires at least 6 months pre-period and 24 months post-period. |
| no\_cont\_eligi\_fu | 1419 | N | **Mean (sd):** 1419 (586), **Min:**-2017, **Max:**3102  There is negative value for number of follow-up days. It should be some missing data. |
| no\_cont\_eligi\_pre\_period | 920 | N | **Mean (sd):** 920 (679), **Min:** 181, **Max:**4197  It is not consistent with the eligibility days as the smallest number of eligibility days is 28 days but the smallest number of continuous eligibility for pre-period is 181 days. |
| insurance | 1-6 | Y | [1 = Commercial (83.79%), 2 = SCHIP (0.53%), 3 = Medicaid (4.02%), 4 = Medicare (0.46%), 5 = Self-insured (10.67%), 6 = Unknown/missing (0.53%)] |
| no\_dx | 2 | N | **Mean (sd):** 5.52 (6.12), **Min:** 0, **Max:** 79  There is 19.71% of zero diagnosis in pre-period. Does it make sense? |
| allergic\_rhinitis | 0/1 | Y | 15.30% |
| obesity | 0/1 | Y | 2.70% |
| eczema | 0/1 | Y | 3.18% |
| osa | 0/1 | Y | 4.17% |
| gerd | 0/1 | Y | 0.34% |
| resp\_infect | 0/1 | Y | 24.60% |
| no\_hosp | 2 | N | **Mean (sd):** 0.025 (1.915), **Min:** 0, **Max:** 10  97.89% of zero hospitalization |
| no\_asthma\_hosp | 4 | N | **Mean (sd):** 0.003 (0.060), **Min:** 0, **Max:** 4  99.73% of zero asthma hospitalization |
| no\_ed | 1 | N | **Mean (sd):** 0.114 (0.317), **Min:** 0, **Max:** 1  88.64% of zero ED visit |
| no\_asthma\_ed | 1 | N | **Mean (sd):** 0.007 (0.082), **Min:** 0, **Max:** 1  99.32% of zero asthma ED visit |
| no\_med\_visit | 20 | N | **Mean (sd):** 4.78 (6.73), **Min:** 0, **Max:** 179  Is it possible to have more than 100 medical visits within 6 months? |
| no\_procedure | 10 | N | **Mean (sd):** 8.04 (10.22), **Min:** 0, **Max:** 142  Is it possible to have more than 100 procedure claims within 6 months? |
| no\_drug\_claim | 3 | Y | **Mean (sd):** 3.38 (4.21), **Min:** 0, **Max:** 54 |
| baseline\_cost | 1000 | N | **Mean (sd):** 1,969 (7,946), **Min:** -8,222, **Max:** 806,192  There are 12.16% of overall observations which had ≤0 cost. Is it possible?  There are nine observation with negative value. |
| no\_asthma\_op | 0-8 | N | **Mean (sd):** 0.017 (0.174), **Min:** 0, **Max:** 8  There is 98.74% of all observations which had 0 asthma op visit. |
| no\_asthma\_lrti | 0-8 | N | **Mean (sd):** 0.026 (0.214), **Min:** 0, **Max:** 8  There is 98.07% of all observations which had 0 lrti visit. |
| CCI | 0-12 | Y | **Mean (sd):** 0.201 (0.576), **Min:** 0, **Max:** 12 |

1. **Exposure**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Format** | **Assessment**  **(Correct? Y/N)** | **Note** |
| pat\_id | 0268AAAAAAAACPBG | N | N = 59,968  Number of observation = 699,454  It is not consistent with N of baseline characteristic file. ***NEED to check*** |
| index\_asthma\_date | 01jan2006 | N | **Min**: 01jan2006, **Max**: 29jun2012  Why are there index date before 01Jul2006 because we need at least 6 months eligibility before the index date? ***(Inclusion criteria#2)***  ***Consistent with baseline characteristic file*** |
| elig\_start | 02jan2006 | Y | **Min**: 01jan2001, **Max**: 01dec2011  ***Consistent with baseline characteristic file*** |
| elig\_stop | 02jan2006 | N | **Min**: 31jan2008, **Max**: 30jun2014  ***NOT consistent with baseline characteristic file*** (Min is 31 jan 2006 for baseline but 31jan2008 for exposure) |
| step\_start | 01jan2006 | N | **Min**: 01jan2006, **Max**: 18jun2014  Out of our observational period (2006-2013) |
| step\_stop | 03jan2006 | N | **Min**: 03jan2006, **Max**: 25nov2021  There are some date stop (2.4%) which are later than 31dec2013 (last date of our observational period) |
| trt\_step | 0-5 | Y | trt0=43.18% ***(based on at least one day gap)***  trt1=17.70%  trt2=14.51%  trt3=11.86%  trt4=11.49%  trt5=1.26% |
| start\_date | 2542 | Y |  |
| stop\_date | 1564 | Y |  |

1. **Outcomes**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Format** | **Assessment**  **(Correct? Y/N)** | **Note** |
| pat\_id | 0268AAAAAAAACPBG | Y | N = 60,479 |
| eventdt | 02jan2006 | Y | **Min**: 01jan2006, **Max**: 31dec2013 |
| event | 0-4 | Y | 0 = 7.74%  1 = 10.83%  2 = 18.37%  3 = 24.78%  4 = 38.27%  **Number of event = 168,746 events** |
| event\_date | 1231 | Y | Attached to eventdt and index date |