

Deliverable 2 – Health Analytics Project Design

Hui Xia
hxia40@gatech.edu

CS-6440 – Introduction to Health Informatics

Georgia Institute of Technology

Feb 16th, 2020

CLINICAL QUESTION

Do patients that have been treated with Anagrelide have a higher risk of cancer, compared with a more conventional alternative, Hydroxyurea?

Note that in Deliverable 1, the clinical question was:

In patients undergoing treatment for thrombocythemia, does those that have been treated with Anagrelide has a higher risk of thrombosis, compared with a more conventional alternative, Hydroxyurea? However, the total number of concept records for such clinical question is limited. Thus, another valid clinical question is needed for this deliverable.

One of the driving force for the development of Anagrelide is that Hydroxyurea has been reported for being related with cancer risk. For example, Nand et. al. reported that hydroxyurea has 1–5.9% risk of causing leukemic transformation (Nand, Stock, Godwin, & Fisher, 1996). Hanft et. al. further suggested that in vivo hydroxyurea exposure could cause acquired DNA mutations (Hanft et al., 2000). Thus, whether such suggestions against hydroxyurea is reasonable, and whether the alternatives thereof, such as Anagrelide, could perform better in cancer risk, will need to be investigated.

PATIENT COUNTS

- [hxia40] patients taking Anagrelide
<http://gt-health-analytics-1.us-east-1.elasticbeanstalk.com/#/cohortdefinition/446>
- [hxia40] patients taking Hydroxyurea
<http://gt-health-analytics-1.us-east-1.elasticbeanstalk.com/#/cohortdefinition/449>
- [hxia40] cancer patients exposed to either drug
<http://gt-health-analytics-1.us-east-1.elasticbeanstalk.com/#/cohortdefinition/452>

Patient counts in both datasets:

Cohort	Patients taking Anagrelide	Patients taking Hydroxyurea	Cancer patients exposed to either drug
CMSDESynPUF100k	638	699	172
CMSDESynPUF23m	20,392	15,774	4,801

INCIDENCE RATES

Among patients taking Anagrelide, 125.39/1,000 patients in the 100k dataset and 129.18/1000 patients in the 23m dataset were diagnosed with cancer. The incidence rates per 1k years are 103.49 and 102.95. Among patients taking Hydroxyurea, 133.05/1000 in the 100k dataset and 138.92/1000 patients in the 23m dataset were diagnosed with cancer. The incidence rates per thousand years are 103.79 and 108.68. Results are shown as below:

Incidence Rate Analysis #73
[hxia40] Anagrelide vs Hydroxyurea

Showing target cohort: [hxia40] patients taking Anagrelide and outcome cohort: [hxia40] cancer patients exposed to either

	Source Name	Persons	Cases	Proportion [+/-] per 1k persons	Time At Risk (years)	Rate [+/-] per 1k years	Started	Duration	
Run	CMSDESynPUF100k	638	80	125.39	773	103.49	02/20/2020 11:46 AM	00:00:16	Reports
Run	CMSDESynPUF23m	20,374	2,632	129.18	25,567	102.95	02/20/2020 11:46 AM	00:00:23	Reports

Incidence Rate Analysis #73
[hxia40] Anagrelide vs Hydroxyurea

Showing target cohort: [hxia40] patients taking Hydroxyurea and outcome cohort: [hxia40] cancer patients exposed to either

	Source Name	Persons	Cases	Proportion [+/-] per 1k persons	Time At Risk (years)	Rate [+/-] per 1k years	Started	Duration	
Run	CMSDESynPUF100k	699	93	133.05	896	103.79	02/20/2020 11:46 AM	00:00:16	Reports
Run	CMSDESynPUF23m	15,764	2,190	138.92	20,150	108.68	02/20/2020 11:46 AM	00:00:23	Reports

<http://gt-health-analytics-1.us-east-1.elasticbeanstalk.com/#/iranalysis/73>

COHORT CHARACTERIZATION

<http://gt-health-analytics-1.us-east-1.elasticbeanstalk.com/#/cc/characterizations/65/design>

STRETCH POINTS:

ATLAS

- Home
- Data Sources
- Search
- Concept Sets
- Cohort Definitions
- Characterizations
- Cohort Pathways
- Incidence Rates
- Profiles
- Estimation
- Prediction
- Jobs
- Configuration
- Feedback

Comparison
Add or update the target, comparator, outcome(s) cohorts and negative control outcomes

Choose your target cohort: [hxia40] patients taking Anagrelide **T**

Choose your comparator cohort: [hxia40] patients taking Hydroxyurea **C**

Choose your outcome cohorts: **Add Outcome**

Show 10 entries

ID	Name	
452	[hxia40] cancer patients exposed to either drug	O

Showing 1 to 1 of 1 entries

Choose your negative control outcomes: [Project] Negative Control Set **Negative control**

Covariate selection

Please note: If you would like to include/exclude covariates based on descendant concepts, it is most efficient to specify this as part of the analysis settings. If you plan to include/exclude descendants, define your concept sets utilizing the ancestor concepts only.

What concepts do you want to include in baseline covariates in the propensity score model? (Leave blank if you want to include everything)

What concepts do you want to exclude in baseline covariates in the propensity score model?

[hxia40] covariate exclusion **Covariate Exclusion, link is provided below**

Apache 2.0
open source software
provided by
OHDSI
join the journey

[hxia40] covariate exclusion:

<http://gt-health-analytics-1.us-east-1.elasticbeanstalk.com/#/conceptset/799/conceptset-expression>

Time At Risk

Define the time-at-risk window start, relative to target/comparator cohort entry:

0 days from cohort start date

Define the time-at-risk window end:

1095 days from cohort end date

The minimum number of days at risk?

1

Study Population

Study start date - a calendar date specifying the minimum date that a cohort index can appear (leave blank to use all time):

YYYY-MM-DD

Study end date - a calendar date specifying the maximum date that a cohort index can appear (leave blank to use all time). **Important:** the study end date is also

YYYY-MM-DD

Should only the first exposure per subject be included?

No

Remove subjects that are in both the target and comparator cohort?

Keep All

Restrict the analysis to the period when both exposures are observed?

No

The minimum required continuous observation time prior to index date for a person to be included in the cohort.

0

If either the target or the comparator cohort is larger than this number it will be sampled to this size. (0 for this value indicates no maximum size)

0

Remove subjects that have the outcome prior to the risk window start?

Yes Remove subjects with prior events

How many days should we look back when identifying prior outcomes?

99999

If a subject is in multiple cohorts, should time-at-risk be censored when the new time-at-risk start to prevent overlap?

No

Below is the result I got from estimation study analysis in RStudio. The patient counts for target and comparator cohorts are 634 and 691. The incidence rates are 103.15 and 103.56, accordingly. There numbers are similar to what I got by estimation analysis at Atlas.

Evidence Explorer

Target

[hxia40] patients taking Anagrelide

Comparator

[hxia40] patients taking Hydroxyurea

Outcome

[hxia40] cancer patients exposed to either drug

Data source

☒ SynpuF

Analysis

☒ New analysis 1

Show 15 entries

Analysis	Data source	HR	LB	UB	P	Cal.HR	Cal.LB	Cal.UB	Cal.P
New analysis 1	SynpuF	0.96	0.71	1.31	0.81	NA	NA	NA	NA

Showing 1 to 1 of 1 entries

Previous 1 Next

Power Attrition Population characteristics Propensity model Propensity scores Covariate balance Systematic error

Table 1a. Number of subjects, follow-up time (in years), number of outcome events, and event incidence rate (IR) per 1,000 patient years (PY) in the target (*[hxia40] patients taking Anagrelide*) and comparator (*[hxia40] patients taking Hydroxyurea*) group after propensity score adjustment, as well as the minimum detectable relative risk (MDRR). Note that the IR does not account for any stratification.

Target subjects	Comparator subjects	Target years	Comparator years	Target events	Comparator events	Target IR (per 1,000 PY)	Comparator IR (per 1,000 PY)	MDRR
634	691	775	898	80	93	103.15	103.56	1.53

Table 1b. Time (days) at risk distribution expressed as minimum (min), 25th percentile (P25), median, 75th percentile (P75), and maximum (max) in the target (*[hxia40] patients taking Anagrelide*) and comparator (*[hxia40] patients taking Hydroxyurea*) cohort after propensity score adjustment.

Cohort	Min	P10	P25	Median	P75	P90	Max
Target	3	115	211	404	788	850	1,078
Comparator	4	124	230	469	799	874	1,089

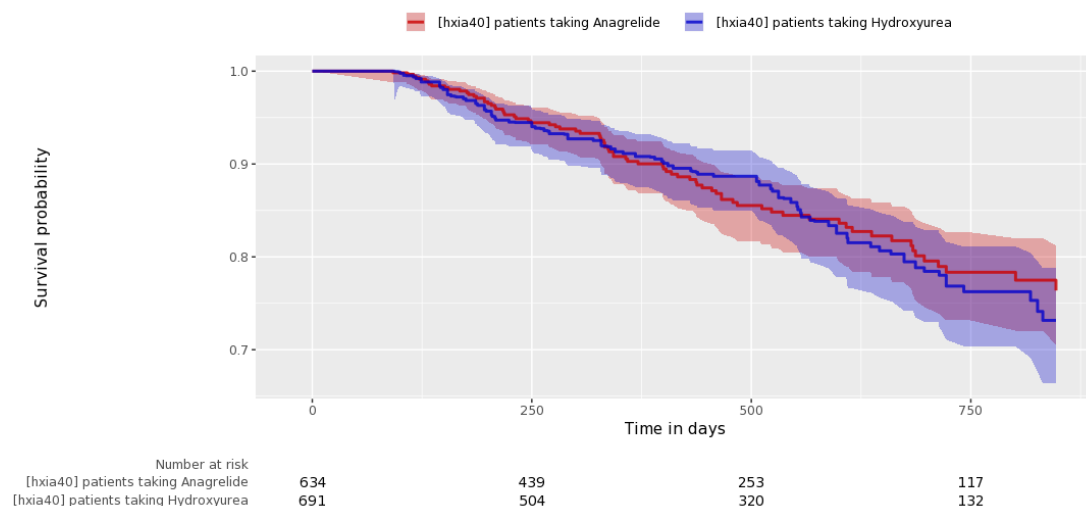


Figure 5. Kaplan Meier plot, showing survival as a function of time. This plot is adjusted using the propensity score: The target curve (*[hxia40] patients taking Anagrelide*) shows the actual observed survival. The comparator curve (*[hxia40] patients taking Hydroxyurea*) applies reweighting to approximate the counterfactual of what the target survival would look like had the target cohort been exposed to the comparator instead. The shaded area denotes the 95 percent confidence interval.

REFERENCES

- 2020 ICD-10-CM Diagnosis Code Z79.02. (2020). Retrieved from <https://www.icd10data.com/ICD10CM/Codes/Z00-Z99/Z77-Z99/Z79-/Z79.02>
- DRUG: Hydroxyurea. Retrieved from https://www.genome.jp/dbget-bin/www_bget?dr:D00341
- Hanft, V. N., Fruchtman, S. R., Pickens, C. V., Rosse, W. F., Howard, T. A., & Ware, R. E. (2000). Acquired DNA mutations associated with in vivo hydroxyurea exposure. *Blood*, 95(11), 3589-3593.
- Nand, S., Stock, W., Godwin, J., & Fisher, S. G. (1996). Leukemogenic risk of hydroxyurea therapy in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *American journal of hematology*, 52(1), 42-46.