Random Frequency-Matching of Controls to Cancer Cases in SEER-Medicare Data by Index Date to Radiation Therapy Date

C. Yee, MS[1,2] , W. Quarshie, MS[1,2] , K. Schwartz, MD, MSPH[1,2]

1) Karmanos Cancer Institute, Detroit, MI, United States 2) Wayne State University, Detroit, MI, United States

BACKGROUND

For a case-control study describing urinary and/or bowel complications post-radiation therapy (RT) in prostate cancer patients

- We needed to randomly match controls to cases
- Not by demographic characteristics
- But by index dates of controls to the RT dates of cancer cases, so that complication claims could be identified for specific periods of time
- We were unable to find a detailed method for this type of matching in the literature

PURPOSE

To demonstrate our method of random frequency-matching of controls to cancer cases by index date to radiation treatment date using the 5% non-cancer random sample of Medicare data.

METHODS

Control selection:

- From the 5% non-cancer random sample of Medicare data
- Include only those who have both Medicare Parts A and B, and no HMO coverage
- For no less than the minimum number of desired follow-up months (MFU), plus 12, counting from 12 months before the earliest radiation therapy month (ERTM)

If MFU = 36

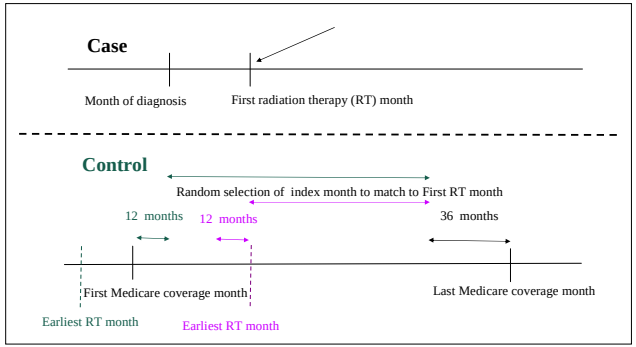
then 36 + 12 = 48 is the minimum Medicare coverage period (with start month ≥ ERTM - 12)

This minimum coverage time is needed to calculate comorbidities 12 months pre- and complications (MFU) post-index date

- Randomly select an index month between the period 12 months after the start of Medicare coverage (or ERTM, whichever is later) and the MFU before the end of coverage.

Cont. METHODS

Based on the frequency counts of cases' RT months by year, we randomly chose the desired number of controls to frequency-match to RT months (4:1). In the subsequent frequency table of controls by index months, if some specific months have a smaller frequency than required, we again performed the random selection of index months for those not selected in the first iteration. We added these to the original sample of controls, and repeated as necessary.



RESULTS

This is one method to produce a sample of randomly matched controls by index dates to RT dates of cancer cases.

Sometimes, besides the random frequency matching of the index dates, it is desirable to proportionately match by another variable , such as age. In our example, the random match resulted in much smaller percentage of controls in the 70-80 age group. So, in the second attempt of random matching, we "adjusted" the second random number to prioritize this particular group so that they are more likely to be selected for the desired frequency matching.

randomN = ranuni(99999); if int(age/10) = 7 then randomN = randomN/4;

4:1 Matching		
Year_Month	Control	Case
200106	12	3
200107	196	49
200108	348	87
200109	532	133
.		
.		

Percentage in Each Age Group			
Age group	Case	ControlSample1	ControlSample2
66 - 69	28.9%	49.1%	35.6%
70 - 74	38.8%	23.6%	34.1%
75 -79	25.8%	14.2%	21.0%
≥ 80	6.5%	13.1%	9.3%

DISCUSSION

Finally, before doing the work of matching non-cancer controls to cases, it should be decided before-hand whether it is the appropriate sample for the case-control analysis. In other words, could the detection bias between the 2 groups be so big to deem the analysis inappropriate?

Sample SAS code:

```
X = INT(ranuni(77777)*(moins-47))+ fstmoins + 11 ;
* Where moins is total # of Medicare months for the potential control
        fstmoins is the 1st month of Medicare insurance (≥ ERTM -12)
        47 because moins-48=0 when moins = 48     => X = 0
IF X = 120 THEN X = 121; * ERTM = 121 (June 2001, in the sumdenom [file of the 5% non-cancer controls] array for insurance coverage);
* Convert back to calendar dates;
moindex = mod(x,12);  * remainder goes from 0 to 11;
yrindex = int(x/12)+1986;
if  moindex = 0 then do;
        yrindex = yrindex - 1;
        moindex = 12;
end;
```