STEPP Subpopulation Analysis for Continuous, Binary and Count Outcomes

Supplementary Material Document

This document contains all the supplementary material associated with the manuscript entitled "STEPP Subpopulation Analysis for Continuous, Binary, and Count Outcomes". It contains the following sections:

1. Complete analyses for the Aspirin/Folate data
   a. Placebo vs. 81 mg. of aspirin
   b. Placebo vs. 325 mg. of aspirin
   c. 81 mg. vs. 325 mg. of aspirin
2. Sensitivity analysis: impact of r_1 and r_2 on Aspirin Study results
3. R Code for the Aspirin/Folate data analyses
4. Analyses using alternative methodologies
   a. Simple logistic regression with an interaction term
   b. MFPI
   c. Virtual Twin Method
5. Description of the New STEPP software
6. Simulation results of null
   a. No treatment effects, no heterogeneity
   b. Constant treatment effects but no heterogeneity

1. Complete analyses for the Aspirin/Folate data

The result of the original study was published in the New England Journal of Medicine [Baron et al 2003] and concluded that low-dose aspirin has a moderate chemopreventive effect on adenomas in the large bowel. We used STEPP to investigate whether the magnitude of the treatment effect is similar across subpopulations defined by patient age. The 3 STEPP analyses are presented below followed by a sensitivity analysis.

STEPP Analysis 1: Placebo vs 81 mg of aspirin

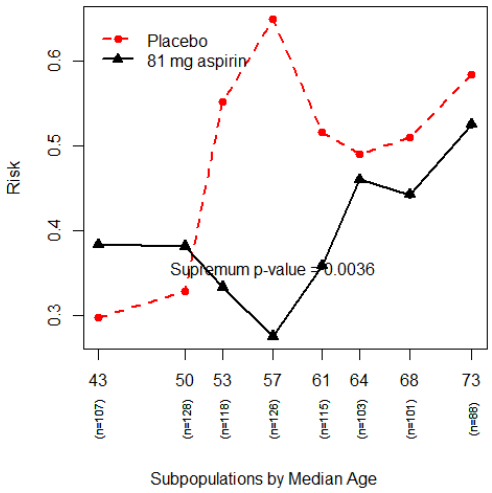
The GLM model for this stepp analysis is

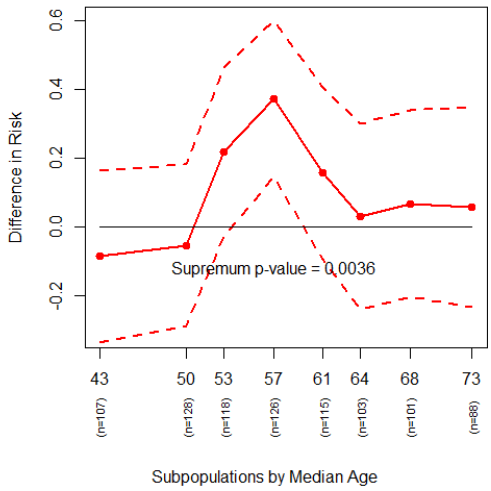
$$logit\,(EY) = \beta 0 + \beta 1 * Treatment$$

The subpopulations are:

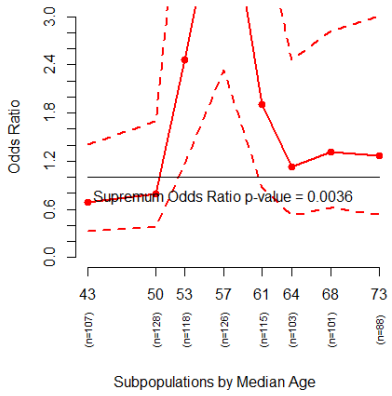
```
        Number of patients per subpopulation (patspop r2): 100
Largest number of patients in common among consecutive subpopulations(minpatspop r1): 30
      Number Of Subpopulations Created : 8

Subpopulation Summary Information
                                Covariate Summary                    Sample
     Subpopulation         Median       Minimum       Maximum          Size
                 1          43.00       29.0000       47.0000           107
                 2          50.00       46.0000       51.0000           128
                 3          53.00       52.0000       55.0000           118
                 4          57.00       55.0000       59.0000           126
                 5          61.00       59.0000       62.0000           115
                 6          64.00       62.0000       66.0000           103
                 7          68.00       66.0000       71.0000           101
                 8          73.00       70.0000       78.0000            88
```

The following are the STEPP plots:





The first STEPP plot shows the risk (or probability of having adenomas) for both treatment groups along different age subgroups – the "red" dashed line is the placebo group and the "black" solid line is the 81mg aspirin group. The 2nd STEPP plot shows the actual differences in risk of getting adenomas in various age subgroups between the placebo and the 81 mg aspirin treatment groups. The interaction p-value based on risk difference is 0.0036 indicating a possible interaction effect between risk and age. It indicates that the effect of the 81 mg to reduce the risk of having adenomas compared with placebo appears to be larger for patients in the middle age subpopulations than it is for either the youngest or oldest subpopulations.



The 3rd STEPP plot shows the odds ratio of getting adenomas in various age subgroups between the placebo and the 81 mg aspirin treatment groups. The overall odds ratio of having adenomas is ~1.46 comparing the placebo vs 81 mg of aspirin treatment groups. The interaction P-value based on odds ratio estimates is also 0.0036 also indicating a possible interaction effect between odds ratio and age.

Looking at these plots, there seems to be a cross over between age 50-64. It is important to note that STEPP simply suggests a hypothesis that there is a reduction of risks of getting adenomas for the 81 mg aspirin group compare with the placebo group for patients between 50 and 64. Further study may be needed to definitively identify the exact cut points.

STEPP Analysis 2: Placebo vs 325 mg of aspirin

The GLM model for this stepp analysis is

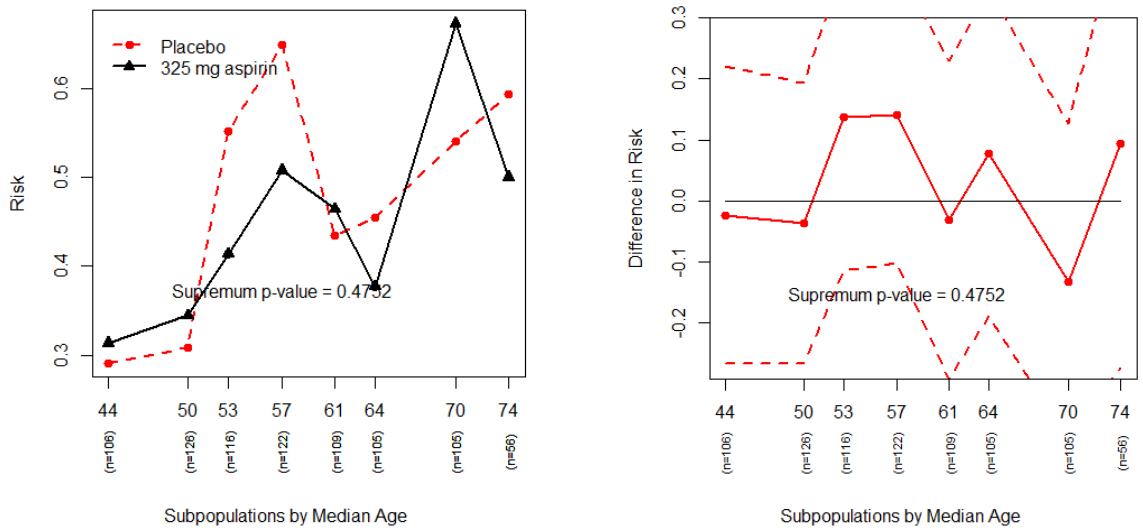
$$logit\,(EY) = \beta 0 + \beta 1 * Treatment$$

with subpopulations of different age groups among treatments of DOSE0 and DOSE2.

The subpopulations are:

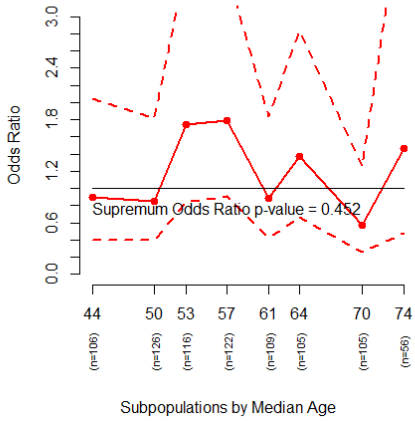
```
        Number of patients per subpopulation (patspop r2): 100
Largest number of patients in common among consecutive subpopulations(minpatspop r1): 30
      Number Of Subpopulations Created : 8

Subpopulation Summary Information
                              Covariate Summary                    Sample
     Subpopulation        Median       Minimum       Maximum         Size
                1             44.00       29.0000       48.0000          106
                2             50.00       47.0000       51.0000          126
                3             53.00       52.0000       55.0000          116
                4             57.00       55.0000       59.0000          122
                5             61.00       60.0000       63.0000          109
                6             64.00       63.0000       67.0000          105
                7             70.00       67.0000       74.0000          105
                8             74.00       72.0000       79.0000           56
```

The following are the STEPP plots:



The first STEPP plot shows the risk (probability of having adenomas) for both treatment groups along different age subgroups – the "red" dashed line is the placebo group and the "black" solid line is the 325mg aspirin group. The 2nd STEPP plot shows the actual differences in risk of getting adenomas in various age subgroups between the placebo and the 325 mg aspirin

treatment groups. The interaction p-value based on risk differences is 0.48 indicating that there may not be any interaction between risk and age.



The 3rd STEPP plot shows the risk ratio of getting adenomas in various age subgroups between the placebo and the 325 mg aspirin treatment groups. The overall odds ratio of having adenomas is ~1.1 comparing the placebo vs 325 mg of aspirin treatment groups. The interaction P-value based on effect ratio is 0.452 also suggesting that there may not be any interaction between risk and age.

There may still be a cross over at age ~50. But the effect is not as drastic in both absolute and relative scale. Furthermore, the p-value is not significant. So, the effect detected could be due to chance.

STEPP Analysis 3: 81 mg of aspirin vs 325 mg of aspirin

The GLM model for this stepp analysis is

$$logit\,(EY) = \beta 0 + \beta 1 * Treatment$$

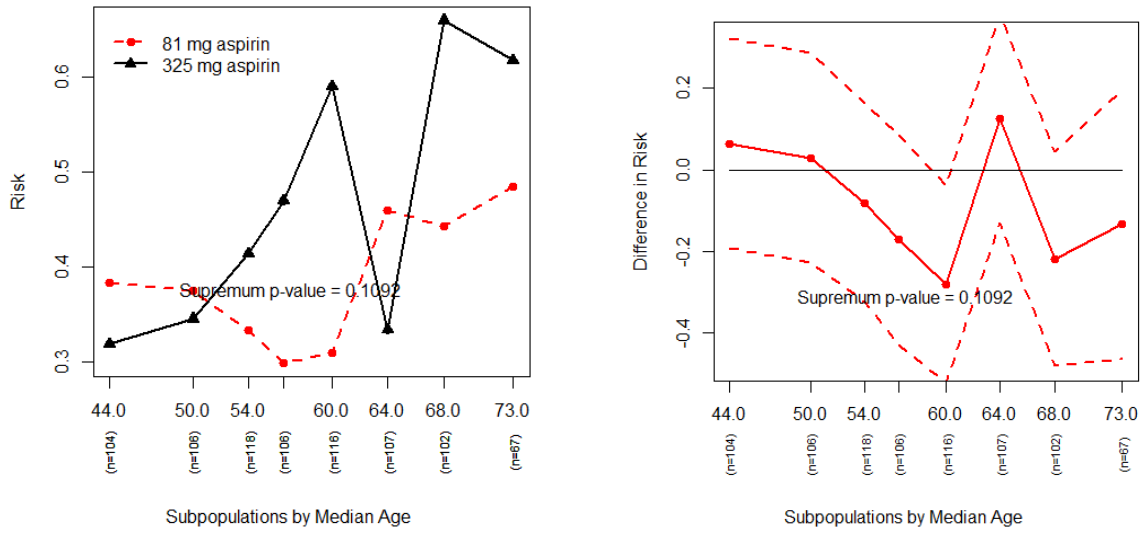
with subpopulations of different age groups among treatments of DOSE1 and DOSE2.

.

The subpopulations are:

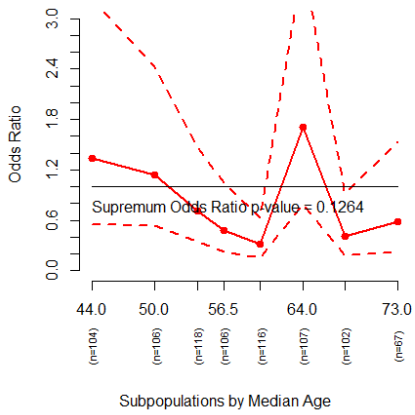
```
        Number of patients per subpopulation (patspop r2): 100
Largest number of patients in common among consecutive subpopulations(minpatspop r1): 30
      Number Of Subpopulations Created : 8

Subpopulation Summary Information
                                Covariate Summary                    Sample
     Subpopulation         Median       Minimum       Maximum          Size
                 1          44.00       29.0000       47.0000           104
                 2          50.00       47.0000       51.0000           106
                 3          54.00       52.0000       55.0000           118
                 4          56.50       55.0000       58.0000           106
                 5          60.00       58.0000       61.0000           116
                 6          64.00       62.0000       66.0000           107
                 7          68.00       66.0000       71.0000           102
                 8          73.00       71.0000       79.0000            67
```

The following are the 3 STEPP plots:



The first STEPP plot shows the risk (probability of having adenomas) for both treatment groups along different age subgroups – the "red" dashed line is the 81 mg aspirin group and the "black" solid line is the 325mg aspirin group. The 2nd STEPP plot shows the actual differences in risk of

getting adenomas in various age subgroups between the 81 mg aspirin and the 325 mg aspirin treatment groups. The interaction P-value based on risk difference is 0.109 indicating a borderline-significant result.



The 3rd STEPP plot shows the odds ratio of getting adenomas in various age subgroups between the 81 mg aspirin and the 325 mg aspirin treatment groups. The overall odds ratio of having adenomas is ~0.75 comparing 81 mg vs 325 mg of aspirin treatment groups. The interaction P-value based on odds ratio is 0.126 indicating a borderline significant interaction effect between treatments and age.

There may be a cross overs at age ~50. The interaction between risk difference and age may be borderline significant in both absolute and relative scale. Looking at the plots, it does seem to have a small beneficial effect for patients taking 81 mg daily vs 325 mg daily between 50 and 60s. Again, STEPP just suggests a hypothesis. Further studies are needed to identify the exact cutoffs.

Conclusion

The results of the aspirin data set confirms with the studies' original result which shows a moderate beneficial effect on getting adenoma with a daily dosage of 81 mg of aspirin. But, in addition, it shows graphically that the risk decreases substantially for patients in age group between 50 and 60. The permutation p value indicates that the interaction is significant. However, the benefit disappears when the daily dosage of 325 mg of aspirin is used. The permutation p value indicates that there does not appear to have any interaction at this dosage. Lastly, it suggests slight risk improvements when comparing the 81 mg daily and 325 mg daily aspirin dosage. The permutation p value indicates only borderline significance of the interaction.

Thus, STEPP confirms the original findings and identifies the age subgroup which may be most benefitted by taking low-dosage aspirin.

2. **Sensitivity Analysis: Impact of r_1 and r_2 on Aspirin Study results**

A sensitivity analysis was done to explore the pattern of change in results when the STEPP smoothing parameters (r_1 and r_2) change. The smoothing parameter r_2, the minimum number of patients in the subpopulation, takes on *100*, *300* or *500* patients out of a total of *725*; we compute r_1, the largest number of patients in common between two consecutive subpopulations, by considering the ratio of r_1/r_2 to be *10%*, *30%*, *50%*, *70%* and *90%*. The number of subpopulations created also changes as values of r_1 and r_2 vary.

There are three tables - one for each comparison scenario. All three STEPP statistics, T_1 (the supremum statistic), T_2 (the chisq statistics in the absolute scale), and T_1* (the supremum statistic in the relative scale) are enumerated.

Table A: Comparing 81 mg of aspirin with placebo. When the ratio r_1/r_2 >= 50%, most of the test statistics become significant. Also, tests are more significant when there are more subpopulations. The highlighted row was used for the aspirin study analysis in the manuscript.

r_2	r_1	r_1/r_2	# of subpop	T_1	T_2	T_1*
100	10	10%	7	0.052	0.093	0.083
	30	30%	8	0.004	0.047	0.011
	50	50%	10	0.006	0.022	0.020
	70	70%	13	0.004	0.008	0.006
	90	90%	21	0.004	0.008	0.006
300	30	10%	3	0.159	0.280	0.236
	90	30%	3	0.041	0.268	0.002
	150	50%	4	<0.001	0.002	<0.001
	210	70%	5	0.038	0.014	0.020
	270	90%	9	0.038	0.014	0.020
500	50	10%	2	0.383	0.588	0.302
	150	30%	2	0.561	0.285	0.361
	250	50%	2	0.037	0.006	0.139
	350	70%	3	0.009	0.010	0.025
	450	90%	4	0.009	0.010	0.025

Table B: Comparing 81 mg of aspirin with 325 mg of aspirin. None of the test statistics is significant for any combination of r_1 and r_2.

r_2	r_1	r_1/r_2	# of	T_1	T_2	T_1*

			subpop			
100	*10*	*10%*	*7*	*0.646*	*0.384*	*0.633*
	30	*30%*	*8*	*0.475*	*0.390*	*0.772*
	50	*50%*	*10*	*0.631*	*0.196*	*0.468*
	70	*70%*	*13*	*0.238*	*0.692*	*0.505*
	90	*90%*	*21*	*0.275*	*0.583*	*0.585*
300	*30*	*10%*	*3*	*0.295*	*0.422*	*0.454*
	90	*30%*	*3*	*0.204*	*0.112*	*0.268*
	150	*50%*	*4*	*0.512*	*0.542*	*0.608*
	210	*70%*	*5*	*0.260*	*0.077*	*0.364*
	270	*90%*	*9*	*0.342*	*0.602*	*0.433*
500	*50*	*10%*	*2*	*0.812*	*0.676*	*0.766*
	150	*30%*	*2*	*0.578*	*0.714*	*0.623*
	250	*50%*	*2*	*0.892*	*0.800*	*0.823*
	350	*70%*	*3*	*0.652*	*0.514*	*0.801*
	450	*90%*	*4*	*0.518*	*0.381*	*0.472*

Table C: Comparing 325 mg of aspirin with placebo. A small number of tests are borderline significant especially when r_1/r_2 is above 70%.

r_2	r_1	r_1/r_2	***# of subpop***	T_1	T_2	T_1*
100	*10*	*10%*	*7*	*0.334*	*0.572*	*0.278*
	30	*30%*	*8*	*0.110*	*0.040*	*0.185*
	50	*50%*	*10*	*0.223*	*0.114*	*0.248*
	70	*70%*	*13*	*0.047*	*0.142*	*0.047*
	90	*90%*	*21*	*0.083*	*0.056*	*0.075*
200	*30*	*10%*	*3*	*0.220*	*0.410*	*0.238*
	90	*30%*	*3*	*0.231*	*0.005*	*0.108*
	150	*50%*	*4*	*0.082*	*0.108*	*0.058*
	210	*70%*	*5*	*0.316*	*0.164*	*0.313*
	270	*90%*	*9*	*0.017*	*0.022*	*0.008*
300	*50*	*10%*	*2*	*0.410*	*0.157*	*0.386*
	150	*30%*	*2*	*0.776*	*0.696*	*0.874*
	250	*50%*	*2*	*0.202*	*0.186*	*0.361*
	350	*70%*	*3*	*0.038*	*0.056*	*0.065*
	450	*90%*	*4*	*0.051*	*0.110*	*0.098*

Comparing 81 mg of aspirin with placebo, the sensitivity analysis result is consistent with our finding that there is evidence in the data to suggest interaction between treatment effect and age. The statistics are nominally significant in most combinations of r_1 and r_2. By contrast, there is no evidence to suggest interaction between treatments of *81* mg of aspirin and *325* mg of aspirin and age. Comparing *325* mg of aspirin with placebo, there are only a few cases that are nominally significant suggesting that evidence supporting an interaction with age is weak.

Note that the result here is specific to this particular clinical trial. Our experience suggests that one cannot make generalizations as the pattern of heterogeneity could be very different in other situations. Similar to smoothing functions, one needs to choose some smoothing parameters for STEPP which may impact the results. For the aspirin trial example, we chose the number of patients per subpopulation *(r_2)* to be 100 and the largest number of patients in common between consecutive subpopulations *(r_1)* to be 30. This choice generates 8 subpopulations providing a good view of the treatment effects along age for analysis. Based on our experience and supported by this sensitivity analysis, the following are general guidelines for choosing r_1 and r_2:

1. Choose r_2 large enough to obtain a good estimate of the treatment effect within subpopulations.
2. Create at least *4-5* subpopulations.
3. Choose r_1/r_2 to be about *30-50%* as your initial investigation.
4. Make r_1, r_2 larger to obtain a smoother STEPP plot, but not so large that you have less than *4* subpopulations.
5. To assess the consistency of the result, a simple sensitivity analysis is recommended by varying r_2.

3. R Code for the Aspirin/Folate data analyses

```
#
# Aspirin Data Analysis
# 9/2/2014
#
library(stepp)
set.seed(1767287)
data(aspirin)
aspirinc <- aspirin[complete.cases(aspirin),]
attach(aspirinc)

# set up the sliding window pattern
inc_win  <- stepp.win(type="sliding", r1=30, r2=100)

# subset the data for appropriate analysis
subset1  <- DOSE == 0  | DOSE == 81
subset2  <- DOSE == 0  | DOSE == 325
subset3  <- DOSE == 81 | DOSE == 325

aspirin1 <- aspirinc[subset1,]
aspirin2 <- aspirinc[subset2,]
aspirin3 <- aspirinc[subset3,]

detach(aspirinc)

trtA <- rep(0, dim(aspirin1)[1])
trtA[aspirin1[,"DOSE"] == 81] <- 1
trtB <- rep(0, dim(aspirin2)[1])
trtB[aspirin2[,"DOSE"] == 325] <- 1
trtC <- rep(0, dim(aspirin3)[1])
trtC[aspirin3[,"DOSE"] == 325] <- 1

#####################################################
# Models for the 3 analysis
#####################################################
#
# STEPP analysis A: placebo vs 81 mg aspirin
attach(aspirin1)
inc_sp  <- stepp.subpop(swin=inc_win, cov=AGE)
summary(inc_win)
summary(inc_sp)

modelA <- stepp.GLM(coltrt=trtA, trts=c(0,1), colY=AD, glm="binomial")
steppGLMA <- stepp.test(inc_sp, modelA, nperm=2500)
summary(steppGLMA)
print(steppGLMA)
plot(steppGLMA, ncex=0.70,legendy=0,
      pline=-4.5, at = 57, color=c("red", "black"),
      xlabel="Subpopulations by Median Age", ylabel="Risk",
      tlegend=c("Placebo", "81 mg aspirin"), nlas=3, pointwise=FALSE,
noyscale=TRUE, rug=FALSE)

detach(aspirin1)
```

```
#
# STEPP analysis B: placebo vs 325 mg aspirin
attach(aspirin2)
inc_sp  <- stepp.subpop(swin=inc_win, cov=AGE)
summary(inc_win)
summary(inc_sp)

modelB <- stepp.GLM(coltrt=trtB, trts=c(0,1), colY=AD, glm="binomial")
steppGLMB <- stepp.test(inc_sp, modelB, nperm=2500)
summary(steppGLMB)
print(steppGLMB)
plot(steppGLMB, ncex=0.70,legendy=0,
      pline=-4.5, at = 57, color=c("red", "black"),
      xlabel="Subpopulations by Median Age", ylabel="Risk",
      tlegend=c("Placebo", "325 mg aspirin"), nlas=3, pointwise=FALSE,
noyscale=TRUE, rug=FALSE)

detach(aspirin2)

#
# STEPP analysis C: 81 mg vs 325 mg of aspirin
attach(aspirin3)
inc_sp  <- stepp.subpop(swin=inc_win, cov=AGE)
summary(inc_win)
summary(inc_sp)

modelC <- stepp.GLM(coltrt=trtC, trts=c(0,1), colY=AD, glm="binomial")
steppGLMC <- stepp.test(inc_sp, modelC, nperm=2500)
summary(steppGLMC)
print(steppGLMC)
plot(steppGLMC, ncex=0.70,legendy=0,
      pline=-4.5, at = 57, color=c("red", "black"),
      xlabel="Subpopulations by Median Age", ylabel="Risk",
      tlegend=c("81 mg aspirin", "325 mg aspirin"), nlas=3, pointwise=FALSE,
noyscale=TRUE, rug=FALSE)

detach(aspirin3)
```

4. Analyses using alternative methodologies

Summary

As a comparison to GLM stepp, we try the following alternative methods to detect interaction between dosage and age in the Aspirin/Folate Polyp Prevention Study data set.

The first method is to model the output with a logistic model with an interaction term (between dosage and age). Then, we can test if the coefficient of the interaction term is significant or not. We treat age first as a continuous covariate and then as a categorical covariate (divided into 3 categories, see below). The coefficients of the interaction term for both of these models are not significant.

We also model the same output using fractional polynomial and use MFPI to assess the interaction. It reports a significant result (with *p*-value = 0.0186) when fp2 is used and patients are grouped into 3 categories with roughly equal number of patients.

Finally, we compare STEPP with a non-parametric method: the Virtual Twin method. Our goals are quite different and so the comparison is not straightforward.

Data

Y denotes the outcome and is equal to 1 if there are adenomas and 0 if none; dose is the dosage (0, 81 and 325 mg of daily aspirin) and is used as treatment (idose is the categorical variable for dose); and age is the age of the patient and is continuous (cage is the categorical variable for age when it is divided into 3 categories: <53, 53-61, >61).

```
. tab cage

       cage |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        344       31.73       31.73
          1 |        375       34.59       66.33
          2 |        365       33.67      100.00
------------+-----------------------------------
      Total |      1,084      100.00

. tab idose

      idose |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        363       33.49       33.49
          1 |        366       33.76       67.25
          2 |        355       32.75      100.00
------------+-----------------------------------
      Total |      1,084      100.00
```

Detail Results from STATA

1. simple logistic regression with an interaction term

a. with an age * idose interaction term

```
. logit y idose age ia

Iteration 0:   log likelihood = -742.04402
Iteration 1:   log likelihood = -730.19729
Iteration 2:   log likelihood = -730.18585
Iteration 3:   log likelihood = -730.18585

Logistic regression                                     Number of obs   =       1084
                                                        LR chi2(3)      =      23.72
                                                        Prob > chi2     =     0.0000
Log likelihood = -730.18585                             Pseudo R2       =     0.0160

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       idose |  -.2193379    .483878    -0.45   0.650    -1.167721    .7290455
         age |   .0286211   .0101471     2.82   0.005     .0087332     .048509
          ia |   .0029839   .0082474     0.36   0.718    -.0131807    .0191485
       _cons |  -1.866695   .5941368    -3.14   0.002    -3.031182    -.702208
------------------------------------------------------------------------------
```

The interaction term, ia, is not statistical significant with *p*-value = 0.718.

b. with an cage * idose interaction term

```
. logit y idose cage iac

Iteration 0:   log likelihood = -742.04402
Iteration 1:   log likelihood = -734.53413
Iteration 2:   log likelihood =   -734.532
Iteration 3:   log likelihood =   -734.532

Logistic regression                                     Number of obs   =       1084
                                                        LR chi2(3)      =      15.02
                                                        Prob > chi2     =     0.0018
Log likelihood =   -734.532                             Pseudo R2       =     0.0101

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       idose |  -.0058596   .1241847    -0.05   0.962    -.2492571    .2375379
        cage |   .3324093   .1187822     2.80   0.005     .0996004    .5652182
         iac |  -.0438832   .0941956    -0.47   0.641    -.2285031    .1407367
       _cons |  -.5548992   .1554056    -3.57   0.000    -.8594886   -.2503099
-------------+----------------------------------------------------------------
```

The interaction term, iac, is also not statistical significant with *p*-value = 0.641.

2. MFPI (Fractional Polynomial)

a. dose as treatment, age as a continuous covariate with logistic regression

i. use fp1

```
. mfpi, treatment(dose) fp1(age) : logistic y
[treating dose as a factor variable, i.dose]

Interactions with i.dose (1084 observations). Flex-1 model (least flexible)

-------------------------------------------------------------------------------
Var         Main        Interact     idf   Chi2     P     Deviance tdf   AIC
-------------------------------------------------------------------------------
age         FP1(.5)     FP1(.5)       2    2.03   0.3630  1452.503  6  1464.503
-------------------------------------------------------------------------------
idf = interaction degrees of freedom; tdf = total model degrees of freedom
```

P-value is 0.363. Using fp1, MFPI does not detect an interaction between age and treatment.

ii. use fp2

```
. mfpi, treatment(dose) fp2(age) : logistic y
[treating dose as a factor variable, i.dose]

Interactions with i.dose (1084 observations). Flex-1 model (least flexible)

-------------------------------------------------------------------------------
Var         Main        Interact     idf   Chi2     P     Deviance tdf   AIC
-------------------------------------------------------------------------------
age         FP2(-2 -2)  FP2(-2 -2)    4    3.92   0.4164  1450.504 10  1470.504
-------------------------------------------------------------------------------
idf = interaction degrees of freedom; tdf = total model degrees of freedom
```

P-value is 0.4164. Using fp2, MFPI does not detect an interaction between age and treatment.

b. dose as treatment, cage (categorized age into 3 groups) with logistic regression

i. use fp1

```
. mfpi, treatment(dose) fp1(cage) : logistic y
[treating dose as a factor variable, i.dose]

Interactions with i.dose (1084 observations). Flex-1 model (least flexible)

-------------------------------------------------------------------------------
Var         Main        Interact     idf   Chi2     P     Deviance tdf   AIC
-------------------------------------------------------------------------------
cage        FP1(-.5)    FP1(-.5)      2    1.63   0.4420  1460.776  6  1472.776
-------------------------------------------------------------------------------
idf = interaction degrees of freedom; tdf = total model degrees of freedom
```

P-value is 0.4420. Using fp1, MFPI does not detect an interaction between cage and treatment.

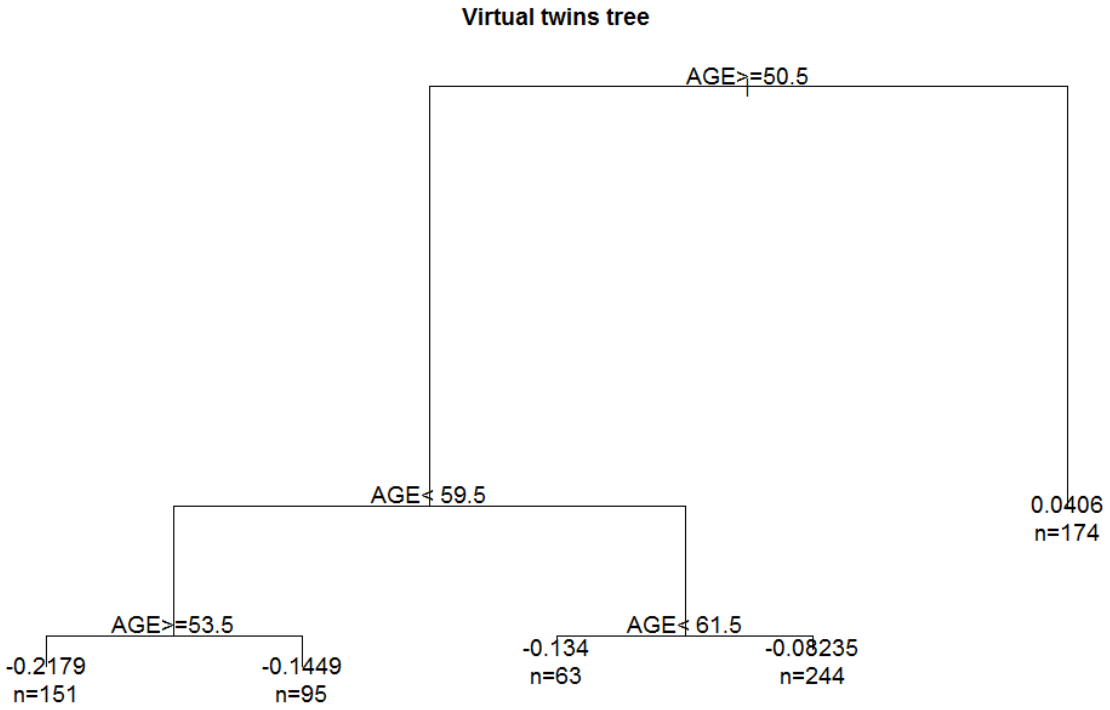
ii. use fp2

```
. mfpi, treatment(dose) fp2(cage) : logistic y
[treating dose as a factor variable, i.dose]

Interactions with i.dose (1084 observations). Flex-1 model (least flexible)

-------------------------------------------------------------------------------
Var         Main        Interact     idf   Chi2     P      Deviance tdf   AIC
-------------------------------------------------------------------------------
cage        FP2(-2 -.5) FP2(-2 -.5)   4   11.83   0.0186  1450.573 10  1470.573
-------------------------------------------------------------------------------
idf = interaction degrees of freedom; tdf = total model degrees of freedom
```

P-value is 0.0186. Using fp2, MFPI detects an interaction between cage and treatment.

3. Virtual Twin Method

We downloaded the Virtual Twin Method software "9-30-13 VT updated for website" to perform a qualitative comparison with STEPP. We apply the Virtual Twin Method to the Aspirin data set to identify the regions of interaction with treatments of Placebo and 81 mg of aspirin with AGE as the only covariate resulting in the following tree:



The Virtual Twin method identifies a similar region of AGE, A={50.5 < AGE < 59.5}, for a potential treatment effect interactions, but the treatment enhancement evaluation, Q(A), was low.

5. Description of the New STEPP software

In a concerted effort, the existing STEPP R package is updated to handle the three GLM models. The latest version of STEPP (version 3.x) is available through CRAN. We redesign a whole new interface so that STEPP analysis for all different models can be done consistently. The old STEPP functions are still being maintained but may be deprecated in the future.

The following are the new interface provided to do a stepp analysis and is implemented as S4 objects:

`stepp.win` – to create a stepp window (**`stwin`**) with r_1 and r_2 as parameters

`stepp.subpop` – to create a stepp subpopulation (**`stsubpop`**) object; to fill in the subpopulation, use the generate method.

`stepp.CI, stepp.KM, stepp.COX, stepp.GLM` – constructor functions to create the corresponding **S4** stepp models: **`stmodelCI, stmodelKM, stmodelCOX and stmodelGLM`**

The **`summary, print and plot`** methods of each of the model generate the resulting tables and 3 stepp plots for analysis.

`stepp.estimate` – apply the stepp model to the subpopulations and estimate their effects.

`stepp.test` – apply permutation and parametric tests to detect the null hypothesis of no heterogeneity among the subpopulations. It produces all the different kinds of estimates, variance covariance matrices and pvalues.

Two data sets are provided:

`aspirin` – aspirin study by John Baron et al.

`big` – big breast cancer study.

For backward compatibility with previous versions of **STEPP**, the old interfaces are maintained:

`analyze.CumInc.stepp`

`analyze.KM.stepp`

`stepp`

`stepp_summary`

`stepp_print`

`stepp_plot`

To see how to use these S4 objects and functions, please refer to the reference manual.

6. Simulation results of null

a. No treatment effects, no heterogeneity

Gaussian	5000 simulations																
Z	N(25,100)																
Y	N(55,49)		100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
		0.100	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
	t1	0.098	0.049	0.008	t1	0.101	0.049	0.009	t1	0.102	0.048	0.007	t1	0.095	0.046	0.007	
	t2	0.106	0.054	0.009	t2	0.105	0.051	0.010	t2	0.101	0.049	0.010	t2	0.1006	0.0526	0.013	
	t1*	0.101	0.048	0.007	t1*	0.101	0.050	0.009	t1*	0.100	0.048	0.007	t1*	0.095	0.046	0.008	
Y	N(75,25)		100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
		0.100	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
	t1	0.099	0.053	0.011	t1	0.108	0.052	0.012	t1	0.108	0.054	0.010	t1	0.099	0.052	0.011	
	t2	0.106	0.058	0.014	t2	0.107	0.055	0.012	t2	0.109	0.056	0.012	t2	0.106	0.053	0.010	
	t1*	0.099	0.052	0.011	t1*	0.107	0.053	0.012	t1*	0.108	0.054	0.009	t1*	0.100	0.052	0.010	
Y	N(95,36)		100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
		0.100	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
	t1	0.096	0.048	0.010	t1	0.101	0.052	0.010	t1	0.102	0.053	0.009	t1	0.104	0.057	0.011	
	t2	0.105	0.051	0.009	t2	0.106	0.053	0.009	t2	0.098	0.049	0.012	t2	0.113	0.057	0.010	
	t1*	0.096	0.048	0.010	t1*	0.099	0.052	0.011	t1*	0.102	0.053	0.009	t1*	0.104	0.057	0.011	

Bernoulli	5000 simulations																
Z	N(25,100)																
Y	Bin(n, 0.3)		100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
		0.100	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
	t1	0.108	0.050	0.009	t1	0.101	0.051	0.011	t1	0.104	0.053	0.010	t1	0.093	0.048	0.011	
	t2	0.102	0.052	0.010	t2	0.107	0.054	0.013	t2	0.110	0.057	0.012	t2	0.105	0.564	0.012	
	t1*	0.100	0.051	0.012	t1*	0.102	0.052	0.011	t1*	0.103	0.051	0.011	t1*	0.096	0.048	0.010	
Y	Bin(n, 0.5)		100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
		0.100	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
	t1	0.094	0.046	0.009	t1	0.102	0.051	0.012	t1	0.104	0.055	0.010	t1	0.099	0.048	0.009	
	t2	0.090	0.045	0.010	t2	0.105	0.053	0.010	t2	0.111	0.057	0.013	t2	0.106	0.052	0.008	
	t1*	0.095	0.047	0.008	t1*	0.103	0.052	0.010	t1*	0.107	0.055	0.010	t1*	0.098	0.048	0.009	
Y	Bin(n, 0.7)		100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
		0.100	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
	t1	0.096	0.049	0.011	t1	0.094	0.050	0.012	t1	0.105	0.049	0.011	t1	0.099	0.051	0.010	
	t2	0.103	0.055	0.014	t2	0.101	0.053	0.012	t2	0.108	0.055	0.013	t2	0.104	0.055	0.011	
	t1*	0.098	0.050	0.008	t1*	0.099	0.047	0.011	t1*	0.109	0.052	0.011	t1*	0.101	0.052	0.009	

Poisson	5000 simulations																
Z	N(25,100)																
Y	Pois(5)		100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
		0.100	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
	t1	0.093	0.051	0.010	t1	0.098	0.047	0.010	t1	0.104	0.052	0.011	t1	0.098	0.053	0.011	
	t2	0.100	0.050	0.009	t2	0.102	0.047	0.011	t2	0.102	0.053	0.008	t2	0.108	0.056	0.009	
	t1*	0.100	0.050	0.009	t1*	0.097	0.047	0.009	t1*	0.102	0.054	0.010	t1*	0.097	0.052	0.010	
Y	Pois(10)		100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
		0.100	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
	t1	0.102	0.054	0.012	t1	0.101	0.050	0.010	t1	0.106	0.049	0.009	t1	0.104	0.054	0.011	
	t2	0.107	0.053	0.011	t2	0.098	0.052	0.010	t2	0.107	0.055	0.011	t2	0.102	0.050	0.011	
	t1*	0.104	0.055	0.011	t1*	0.103	0.050	0.010	t1*	0.106	0.048	0.009	t1*	0.105	0.053	0.011	
Y	Pois(15)		100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
		0.100	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
	t1	0.100	0.048	0.008	t1	0.102	0.052	0.010	t1	0.094	0.048	0.013	t1	0.105	0.051	0.010	
	t2	0.103	0.052	0.012	t2	0.108	0.054	0.011	t2	0.106	0.057	0.011	t2	0.103	0.052	0.010	
	t1*	0.101	0.048	0.008	t1*	0.104	0.056	0.011	t1*	0.094	0.050	0.012	t1*	0.106	0.052	0.010	

b. Constant treatment effect but no heterogeneity

Gaussian	5000 simulations																	
Z	N(25,100)																	
Y (ctrl)	N(95,36)			100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
Y (trt)	N(95+eff,36)		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
eff	5	t1	0.095	0.046	0.009	t1	0.094	0.050	0.009	t1	0.097	0.051	0.012	t1	0.097	0.048	0.011	t1
		t2	0.103	0.052	0.010	t2	0.102	0.048	0.010	t2	0.110	0.057	0.012	t2	0.100	0.055	0.010	t2
		t1*	0.096	0.047	0.009	t1*	0.096	0.049	0.008	t1*	0.098	0.050	0.011	t1*	0.098	0.049	0.011	t1*

Binomial	5000 simulations																	
Z	N(25,100)																	
Y (ctrl)	Bin(n, 0.5)			100, 30, 40				200, 60, 80				500, 150, 200				1000, 300, 400		
Y (trt)	Bin(n, 0.5 + eff)		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
eff	0.1	t1	0.100	0.051	0.012	t1	0.101	0.053	0.010	t1	0.099	0.048	0.010	t1	0.108	0.056	0.013	t1
		t2	0.104	0.055	0.011	t2	0.106	0.054	0.010	t2	0.104	0.055	0.011	t2	0.109	0.056	0.013	t2
		t1*	0.099	0.049	0.012	t1*	0.100	0.054	0.011	t1*	0.099	0.049	0.010	t1*	0.107	0.055	0.012	t1*

Poisson	5000 simulations																	
Z	N(25,100)																	
Y (ctrl)	Pois(5)			100, 30, 40				100, 60, 80				500, 150, 200				1000, 300, 400		
Y (trt)	Pois(5+eff)		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**		**0.100**	**0.050**	**0.010**	
eff	1	t1	0.105	0.053	0.012	t1	0.096	0.050	0.011	t1	0.102	0.053	0.009	t1	0.099	0.049	0.011	t1
		t2	0.109	0.058	0.011	t2	0.098	0.051	0.010	t2	0.103	0.053	0.011	t2	0.111	0.059	0.012	t2
		t1*	0.106	0.054	0.013	t1*	0.098	0.051	0.010	t1*	0.101	0.054	0.009	t1*	0.099	0.049	0.011	t1*

STEPP Subpopulation Analysis for Continuous, Binary and Count Outcomes

Manuscript ID:	CT-15-0171.R1
Funding Information:	**Italian Ministry of Education, University and Research** × 2007AYHZWC U.S. Department of Health and Human Services > National Institutes of Health > **National Cancer Institute** P30 DE020752 U.S. Department of Health and Human Services > **National Institutes of Health** (2 of 2) COPD Grant/5R01HL089856-08 Lung Disease TG/T32 HL007427
Submitting Author:	Yip, Wai-Ki Save Yip, Wai-Ki (proxy) *primary affiliation* Dana Farber Cancer Institute - Biostatistics and Computation Biology 3 Blackfan Circle 11th floor Boston Massachusetts 02115 United States T: 6176326574
Authors & Institutions:	Yip, Wai-Ki proxy — Dana Farber Cancer Institute - Biostatistics and Computation Biology, 3 Blackfan Circle 11th floor , Boston, Massachusetts 02115, United States Gelber, Richard D proxy — DFCI - Biostatistics and Computational Biology, CLSB 11007 450 Brookline Ave. , Boston, Massachusetts 02215, United States Bonetti, Marco proxy — Milan, Italy Cole, Bernard proxy — Unversity of Vermont, Burlington, Vermont, United States Lazar, Ann proxy — University of California, San Francisco, San Francisco, California, United States Barcella, William proxy — UCL, London, United Kingdom of Great Britain and Northern Ireland Wang, Xin Victoria proxy — 450 Brookline Ave , Boston, Massachusetts 02215, United States
Contact Author (populates the ##PROLE_AUTHOR_..## e-mail tags):	Yip, Wai-Ki Save Current Contact Author: Yip, Wai-Ki (proxy)

Running Head:	GLM STEPP
Keywords:	Randomized Clinicala Trials (RCT), Subpopulation treatmente effect pattern plot (STEPP), Generalized Linear Models (GLM), Subgroup Analysis