

Final Draft, don't edit on this one

Han Cui, Jiaheng Cai

11/19/2021

System information:

```
## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5    v purrr  0.3.4
## v tibble  3.1.5    v dplyr  1.0.7
## v tidyr   1.1.4    v stringr 1.4.0
## v readr   2.1.0    v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

## #BlackLivesMatter

## Loading required package: ggpubr

##
## Attaching package: 'survminer'

## The following object is masked from 'package:survival':
##
##      myeloma

## Learn more about sjPlot with 'browseVignettes("sjPlot")'.
```

Table 1. Baseline Characteristics of WHEL Study Participants by Study Group

```
baseline.data = read_excel(paste0(path, 'demographics.xls'))
baseline.medical.data = read_excel(paste0(path, 'Medical.xls'))
endpoints.data = read_excel(paste0(path, 'endpoints.xls'))

baseline.colname = c('ID', 'Age at Diagnosis', 'Age at Randomization',
                     'Ethnicity', 'Education', 'Marital Status',
                     'Employment Status', 'Year Randomized')
baseline.medical.colname = c('ID', 'Lumpectomy/Mastectomy', 'Radiation',
                             'Chemotherapy', 'Tumor type',
```

```

        'Estrogen Receptor Status',
        'Progesterone Receptor Status',
        'Number axillary lymph nodes examined',
        'Number axillary lymph nodes positive', 'Stage',
        'Tumor size (largest dimension)', 'Height (cm)',
        'Weight (kg)', 'BMI', 'Waist (cm)', 'Hip (cm)',
        'Pulse (30 sec)', 'Blood pressure - systolic',
        'Blood pressure - diastolic', 'Menopause status',
        'Use of Anti-estrogens', 'Type of Anti-estrogen')
colnames(baseline.data) = baseline.colname
colnames(baseline.medical.data) = baseline.medical.colname
baseline.char.data = merge(baseline.data, baseline.medical.data, by.y='ID',
                           all.x=TRUE)

for (i in 1:nrow(baseline.char.data)) {
  baseline.char.data$Ethnicity[i] =
    switch(as.character(baseline.char.data$Ethnicity[i]),
      '1' = 'White, not Hispanic',
      '2' = 'Black not Hispanic',
      '3' = 'Hispanic',
      '4' = 'American Indian/Alaskan',
      '5' = 'Asian',
      '6' = 'Pacific Islander',
      '7' = 'Mixed Race',
      '8' = 'Other')

  baseline.char.data$Education[i] =
    switch(as.character(baseline.char.data$Education[i]),
      '5' = 'High School Graduate or less',
      '6' = 'Post High School Training',
      '7' = 'Some College Education',
      '8' = 'College/University Graduate',
      '9' = 'Post College/University Education')

  baseline.char.data$`Marital Status`[i] =
    switch(as.character(baseline.char.data$`Marital Status`[i]),
      '1' = 'Single',
      '2' = 'Married',
      '3' = 'Separated',
      '4' = 'Divorced',
      '5' = 'Widowed',
      '6' = 'Unknown')

  baseline.char.data$`Employment Status`[i] =
    switch(as.character(baseline.char.data$`Employment Status`[i]),
      '0' = 'Not employed',
      '1' = 'Employed',
      '2' = 'Unknown')

  baseline.char.data$`Lumpectomy/Mastectomy`[i] =
    switch(as.character(baseline.char.data$`Lumpectomy/Mastectomy`[i]),
      '1' = 'Lumpectomy',
      '2' = 'Mastectomy',

```

```

        '3' = 'Neither')

baseline.char.data$Radiation[i] =
  switch(as.character(baseline.char.data$Radiation[i]),
    '1' = 'No',
    '2' = 'Yes',
    '3' = 'Unknown')

baseline.char.data$Chemotherapy[i] =
  switch(as.character(baseline.char.data$Chemotherapy[i]),
    '1' = 'No',
    '2' = 'Yes',
    '3' = 'Unknown')

baseline.char.data$`Tumor type`[i] =
  switch(as.character(baseline.char.data$`Tumor type`[i]),
    '1' = 'Infiltrating ductal and lobular invasive',
    '2' = 'Lobular invasive and not ductal',
    '3' = 'Infiltrating ductal and not lobular invasive',
    '4' = 'No ductal or lobular carcinoma; only medullary, mucoid, papillary,
    tubular, anaplastic, signet ring cell or no tumor found')

baseline.char.data$`Estrogen Receptor Status`[i] =
  switch(as.character(baseline.char.data$`Estrogen Receptor Status`[i]),
    '0' = 'Negative',
    '1' = 'Positive',
    '2' = 'Not Done',
    '3' = 'Unknown')

baseline.char.data$`Progesterone Receptor Status`[i] =
  switch(as.character(baseline.char.data$`Progesterone Receptor Status`[i]),
    '0' = 'Negative',
    '1' = 'Positive',
    '2' = 'Not Done',
    '3' = 'Unknown')

baseline.char.data$`Menopause status`[i] =
  switch(as.character(baseline.char.data$`Menopause status`[i]),
    '1' = 'Premenopausal',
    '2' = 'Postmenopausal',
    '3' = 'Perimenopausal',
    '4' = 'Not sure',
    '5' = 'Other/Unknown')

baseline.char.data$`Use of Anti-estrogens`[i] =
  switch(as.character(baseline.char.data$`Use of Anti-estrogens`[i]),
    '1' = 'Yes, now',
    '2' = 'Never',
    '3' = 'Previously',
    '4' = 'Unknown')

baseline.char.data$`Type of Anti-estrogen`[i] =
  switch(as.character(baseline.char.data$`Type of Anti-estrogen`[i]),

```

```

      '1' = 'Tamoxifen/Nolvadex',
      '2' = 'Raloxifene/Evista',
      '3' = 'Anastrozole/Arimidex',
      '4' = 'None',
      '5' = 'Unknown',
      '6' = 'Letrozole/Femara',
      '7' = 'Toremifene/Faristone')
}

baseline.char.data$`Number axillary lymph nodes examined`[baseline.char.data$`Number axillary lymph nodes examined` == -9] = NA
baseline.char.data$`Number axillary lymph nodes positive`[baseline.char.data$`Number axillary lymph nodes positive` == -9] = NA
baseline.char.data$`Pulse (30 sec)`[baseline.char.data$`Pulse (30 sec)` == -9] = NA
baseline.char.data$`Blood pressure - systolic`[baseline.char.data$`Blood pressure - systolic` == -9] = NA
baseline.char.data$`Blood pressure - diastolic`[baseline.char.data$`Blood pressure - diastolic` == -9] = NA

endpoints.data = read_excel(paste0(path, 'endpoints.xls'))
endpoints_colnames = c('ID', 'Intervention Group',
  'Vitality Status as of 6/1/2006',
  'Breast Cancer Status as of 6/1/2006 or last prior contact',
  'Other Cancer (invasive, not breast) Status as of 6/1/2006',
  'Breast Cancer Contribute to Death',
  'Year Breast Cancer Diagnosed', 'Cancer Grade',
  'Dummy Variable for Cancer Grade 2',
  'Dummy Variable for Cancer Grade 3',
  'Dummy Variable for Unknown Cancer Grade',
  'Cancer Stage, AJCC 6th',
  'Dummy Variable for Stage 2 AJCC 6th',
  'Dummy Variable for Stage 3 AJCC 6th',
  'WHEL Clinical Site', 'Recurrence Flag',
  'Years from Diagnosis to WHEL Study Entry',
  'Years from Study Entry to Recurrence/New Primary, or to Censor',
  'Years from Diagnosis to Recurrence/New Primary, or to Censor',
  'Years from Diagnosis to Death or Censor')

colnames(endpoints.data) = endpoints_colnames
baseline.char.data = cbind(baseline.char.data, endpoints.data$`Intervention Group`)
colnames(baseline.char.data)[ncol(baseline.char.data)] = 'Intervention Group'

baseline.char.data$`Intervention Group` = ifelse(
  baseline.char.data$`Intervention Group` == 3, 'Intervention', 'Comparison')

baseline.char.data %>%
  tbl_summary(by = 'Intervention Group',
    include=-c(ID),
    missing_text = "Missing") %>%
  add_p()

```

```

## There was an error in 'add_p()/add_difference()' for variable 'Ethnicity', p-value omitted:
## Error in stats::fisher.test(c("White, not Hispanic", "White, not Hispanic", : FEXACT error 7(location

```

```
## (pastp=186.409, ipn_0:=ipoin[itp=80]=880, stp[ipn_0]=186.461).
## Increase workspace or consider using 'simulate.p.value=TRUE'
```

```
## Table printed with 'knitr::kable()', not {gt}. Learn why at
## http://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include 'message = FALSE' in code chunk header.
```

Characteristic	Comparison, N = 1,551	Intervention, N = 1,537	p- value
Age at Diagnosis	50 (45, 57)	50 (45, 57)	0.7
Age at Randomization	52 (47, 59)	52 (47, 59)	0.6
Ethnicity			
American Indian/Alaskan	1 (<0.1%)	2 (0.1%)	
Asian	50 (3.2%)	46 (3.0%)	
Black not Hispanic	57 (3.7%)	61 (4.0%)	
Hispanic	78 (5.0%)	87 (5.7%)	
Mixed Race	20 (1.3%)	10 (0.7%)	
Other	7 (0.5%)	12 (0.8%)	
Pacific Islander	10 (0.6%)	13 (0.8%)	
White, not Hispanic	1,328 (86%)	1,306 (85%)	
Education			0.3
College/University Graduate	427 (28%)	454 (30%)	
High School Graduate or less	203 (13%)	176 (11%)	
Post College/University Education	394 (25%)	399 (26%)	
Post High School Training	38 (2.5%)	49 (3.2%)	
Some College Education	489 (32%)	459 (30%)	
Marital Status			0.4
Divorced	199 (13%)	195 (13%)	
Married	1,078 (70%)	1,081 (70%)	
Separated	12 (0.8%)	18 (1.2%)	
Single	177 (11%)	162 (11%)	
Unknown	15 (1.0%)	7 (0.5%)	
Widowed	70 (4.5%)	74 (4.8%)	
Employment Status			0.6
Employed	1,109 (72%)	1,105 (72%)	
Not employed	431 (28%)	425 (28%)	
Unknown	11 (0.7%)	7 (0.5%)	
Year Randomized			>0.9
1995	37 (2.4%)	38 (2.5%)	
1996	225 (15%)	232 (15%)	
1997	256 (17%)	263 (17%)	
1998	329 (21%)	315 (20%)	
1999	376 (24%)	360 (23%)	
2000	328 (21%)	329 (21%)	
Lumpectomy/Mastectomy			0.5
Lumpectomy	750 (48%)	724 (47%)	
Mastectomy	801 (52%)	812 (53%)	
Neither	0 (0%)	1 (<0.1%)	
Radiation			0.5
No	586 (38%)	599 (39%)	
Unknown	3 (0.2%)	1 (<0.1%)	
Yes	962 (62%)	937 (61%)	

Characteristic	Comparison, N = 1,551	Intervention, N = 1,537	p- value
Chemotherapy			0.077
No	487 (31%)	442 (29%)	
Unknown	2 (0.1%)	0 (0%)	
Yes	1,062 (68%)	1,095 (71%)	
Tumor type			0.2
Infiltrating ductal and lobular invasive	83 (5.4%)	81 (5.3%)	
Infiltrating ductal and not lobular invasive	1,270 (82%)	1,220 (79%)	
Lobular invasive and not ductal	124 (8.0%)	150 (9.8%)	
No ductal or lobular carcinoma; only medullary, mucoid, papillary, tubular, anaplastic, signet ring cell or no tumor found	74 (4.8%)	86 (5.6%)	
Estrogen Receptor Status			0.3
Negative	403 (26%)	353 (23%)	
Not Done	8 (0.5%)	7 (0.5%)	
Positive	1,128 (73%)	1,165 (76%)	
Unknown	12 (0.8%)	12 (0.8%)	
Progesterone Receptor Status			0.7
Negative	489 (32%)	497 (32%)	
Not Done	20 (1.3%)	14 (0.9%)	
Positive	1,027 (66%)	1,010 (66%)	
Unknown	15 (1.0%)	16 (1.0%)	
Number axillary lymph nodes examined	14 (10, 18)	14 (11, 19)	0.6
Missing	1	2	
Number axillary lymph nodes positive	0.00 (0.00, 2.00)	0.00 (0.00, 2.00)	0.6
Missing	0	1	
Stage			0.8
I	606 (39%)	585 (38%)	
II	867 (56%)	876 (57%)	
IIIA	78 (5.0%)	76 (4.9%)	
Tumor size (largest dimension)	2.00 (1.40, 2.60)	2.00 (1.35, 2.80)	0.8
Missing	3	2	
Height (cm)	164 (160, 168)	164 (160, 168)	0.7
Weight (kg)	70 (61, 81)	70 (62, 82)	0.5
BMI	25.9 (22.9, 30.1)	26.1 (22.9, 30.3)	0.6
Waist (cm)	83 (74, 94)	83 (75, 94)	0.4
Missing	7	11	
Hip (cm)	104 (98, 112)	105 (98, 112)	0.5
Missing	7	11	
Pulse (30 sec)	34.0 (31.0, 37.0)	34.0 (31.0, 38.0)	0.7
Missing	642	681	
Blood pressure – systolic	115 (106, 126)	115 (106, 125)	0.7
Missing	627	655	
Blood pressure – diastolic	74 (69, 81)	73 (68, 80)	0.14
Missing	628	655	
Menopause status			0.8

Characteristic	Comparison, N = 1,551	Intervention, N = 1,537	p- value
Not sure	2 (0.1%)	3 (0.2%)	
Perimenopausal	145 (9.3%)	140 (9.1%)	
Postmenopausal	1,220 (79%)	1,228 (80%)	
Premenopausal	184 (12%)	166 (11%)	
Use of Anti-estrogens			0.026
Never	525 (34%)	451 (29%)	
Previously	112 (7.2%)	107 (7.0%)	
Unknown	1 (<0.1%)	3 (0.2%)	
Yes, now	913 (59%)	976 (64%)	
Type of Anti-estrogen			0.046
Anastrozole/Arimidex	1 (<0.1%)	4 (0.3%)	
None	525 (34%)	451 (29%)	
Raloxifene/Evista	17 (1.1%)	19 (1.2%)	
Tamoxifen/Nolvadex	992 (64%)	1,046 (68%)	
Toremifene/Faristone	0 (0%)	2 (0.1%)	
Unknown	16 (1.0%)	15 (1.0%)	

Data clean and recode

```

endpoint = read_excel(paste0(path, 'endpoints.xls'))
endpoint_colnames = c('ID', 'Intervention Group',
                      'Vitality Status as of 6/1/2006',
                      'Breast Cancer Status as of 6/1/2006 or last prior contact',
                      'Other Cancer (invasive, not breast) Status as of 6/1/2006',
                      'Breast Cancer Contribute to Death',
                      'Year Breast Cancer Diagnosed', 'Cancer Grade',
                      'Dummy Variable for Cancer Grade 2',
                      'Dummy Variable for Cancer Grade 3',
                      'Dummy Variable for Unknown Cancer Grade',
                      'Cancer Stage, AJCC 6th',
                      'Dummy Variable for Stage 2 AJCC 6th',
                      'Dummy Variable for Stage 3 AJCC 6th',
                      'WHEL Clinical Site', 'Recurrence Flag',
                      'Years from Diagnosis to WHEL Study Entry',
                      'Years from Study Entry to Recurrence/New Primary, or to Censor',
                      'Years from Diagnosis to Recurrence/New Primary, or to Censor',
                      'Years from Diagnosis to Death or Censor')

colnames(endpoint) = endpoint_colnames

endpoint$`Intervention Group` = ifelse(endpoint$`Intervention Group` == 3,
                                       'Intervention', 'Comparison')

endpoint$`Vitality Status as of 6/1/2006` =
  ifelse(endpoint$`Vitality Status as of 6/1/2006` == 0, 'Dead',
        ifelse(endpoint$`Vitality Status as of 6/1/2006` == 1, 'Alive', 'Unknown'))

for (i in 1:nrow(endpoint)){
  if (endpoint$`Breast Cancer Status as of 6/1/2006 or last prior contact`[i] == 0){
    endpoint$`Breast Cancer Status as of 6/1/2006 or last prior contact`[i] =
      'No Evidence of Recurrence'
  }
}

```

```

} else if(endpoint$`Breast Cancer Status as of 6/1/2006 or last prior contact`[i] == 1){
  endpoint$`Breast Cancer Status as of 6/1/2006 or last prior contact`[i] =
    'Confirmed - New Primary Breast Cancer'
} else if(endpoint$`Breast Cancer Status as of 6/1/2006 or last prior contact`[i] == 2){
  endpoint$`Breast Cancer Status as of 6/1/2006 or last prior contact`[i] =
    'Confirmed - Local Recurrence'
} else if(endpoint$`Breast Cancer Status as of 6/1/2006 or last prior contact`[i] == 3){
  endpoint$`Breast Cancer Status as of 6/1/2006 or last prior contact`[i] =
    'Confirmed - Regional Recurrence'
} else{
  endpoint$`Breast Cancer Status as of 6/1/2006 or last prior contact`[i] =
    'Confirmed - Distant Recurrence '
}

endpoint$`Other Cancer (invasive, not breast) Status as of 6/1/2006` =
  ifelse(endpoint$`Other Cancer (invasive, not breast) Status as of 6/1/2006`
    == 0, 'No evidence of Disease',
    ifelse(endpoint$`Other Cancer (invasive, not breast) Status as of 6/1/2006` == 1,
      'Reported Other Cancer (not confirmed)', 'Confirmed Other Cancer'))

endpoint$`Breast Cancer Contribute to Death`[endpoint$`Breast Cancer Contribute to Death` == -1] =
  'Not Dead'
endpoint$`Breast Cancer Contribute to Death`[endpoint$`Breast Cancer Contribute to Death` == 0] =
  'Dead from a cause other than Breast Cancer'
endpoint$`Breast Cancer Contribute to Death`[endpoint$`Breast Cancer Contribute to Death` == 1] =
  'Dead from Breast Cancer'
endpoint$`Breast Cancer Contribute to Death`[endpoint$`Breast Cancer Contribute to Death` == 2] =
  'Dead from Cancer, not confirmed breast but likely so'

endpoint$`Cancer Grade`[endpoint$`Cancer Grade` == 0] = 'Grade Not Applicable or Not Available'
endpoint$`Cancer Grade`[endpoint$`Cancer Grade` == 1] = 'Grade I, Well Differentiated'
endpoint$`Cancer Grade`[endpoint$`Cancer Grade` == 2] = 'Grade II, Moderately Differentiated'
endpoint$`Cancer Grade`[endpoint$`Cancer Grade` == 3] = 'Grade III, Poorly Differentiated'

endpoint$`Cancer Stage, AJCC 6th`[endpoint$`Cancer Stage, AJCC 6th` == 1] = 'Stage I'
endpoint$`Cancer Stage, AJCC 6th`[endpoint$`Cancer Stage, AJCC 6th` == 2] = 'Stage IIA'
endpoint$`Cancer Stage, AJCC 6th`[endpoint$`Cancer Stage, AJCC 6th` == 3] = 'Stage IIB'
endpoint$`Cancer Stage, AJCC 6th`[endpoint$`Cancer Stage, AJCC 6th` == 4] = 'Stage IIIA'
endpoint$`Cancer Stage, AJCC 6th`[endpoint$`Cancer Stage, AJCC 6th` == 5] = 'Stage IIIB'

endpoint$`WHEL Clinical Site`[endpoint$`WHEL Clinical Site` == 1] = 'Site A in California'
endpoint$`WHEL Clinical Site`[endpoint$`WHEL Clinical Site` == 3] = 'Site B in California'
endpoint$`WHEL Clinical Site`[endpoint$`WHEL Clinical Site` == 5] = 'Site C in California'
endpoint$`WHEL Clinical Site`[endpoint$`WHEL Clinical Site` == 7] = 'Site in Arizona'
endpoint$`WHEL Clinical Site`[endpoint$`WHEL Clinical Site` == 9] = 'Site D in California'
endpoint$`WHEL Clinical Site`[endpoint$`WHEL Clinical Site` == 11] = 'Site in Texas'
endpoint$`WHEL Clinical Site`[endpoint$`WHEL Clinical Site` == 13] = 'Site in Oregon'

endpoint$`Recurrence Flag`[endpoint$`Recurrence Flag` == 0] = 'No Invasive Breast Cancer Events'
endpoint$`Recurrence Flag`[endpoint$`Recurrence Flag` == 1] = 'Invasive Breast Cancer Event'

```


Baseline Characteristics of WHEL Study Participants by Study Group

```
endpoint %>%
  tbl_summary(by = 'Intervention Group',
              include=c(ID, `Year Breast Cancer Diagnosed`,
                          `Dummy Variable for Cancer Grade 2`,
                          `Dummy Variable for Cancer Grade 3`,
                          `Dummy Variable for Unknown Cancer Grade`,
                          `Dummy Variable for Stage 2 AJCC 6th`,
                          `Dummy Variable for Stage 3 AJCC 6th`),
              missing_text = "Missing") %>%
  add_p()
```

Table printed with 'knitr::kable()', not {gt}. Learn why at
 ## <http://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html>
 ## To suppress this message, include 'message = FALSE' in code chunk header.

Characteristic	Comparison, N = 1,551	Intervention, N = 1,537	p-value
Vitality Status as of 6/1/2006			0.8
Alive	1,391 (90%)	1,381 (90%)	
Dead	160 (10%)	155 (10%)	
Unknown	0 (0%)	1 (<0.1%)	
Breast Cancer Status as of 6/1/2006 or last prior contact			0.6
Confirmed – Distant Recurrence	189 (12%)	168 (11%)	
Confirmed – Local Recurrence	28 (1.8%)	35 (2.3%)	
Confirmed – New Primary Breast Cancer	35 (2.3%)	43 (2.8%)	
Confirmed – Regional Recurrence	10 (0.6%)	10 (0.7%)	
No Evidence of Recurrence	1,289 (83%)	1,281 (83%)	
Other Cancer (invasive, not breast) Status as of 6/1/2006			0.5
Confirmed Other Cancer	60 (3.9%)	58 (3.8%)	
No evidence of Disease	1,483 (96%)	1,472 (96%)	
Reported Other Cancer (not confirmed)	4 (0.3%)	1 (<0.1%)	
Missing	4	6	
Breast Cancer Contribute to Death			0.8
Dead from a cause other than Breast Cancer	25 (1.6%)	28 (1.8%)	
Dead from Breast Cancer	135 (8.7%)	126 (8.2%)	
Dead from Cancer, not confirmed breast but likely so	0 (0%)	1 (<0.1%)	
Not Dead	1,391 (90%)	1,382 (90%)	
Cancer Grade			>0.9
Grade I, Well Differentiated	245 (16%)	239 (16%)	
Grade II, Moderately Differentiated	620 (40%)	620 (40%)	
Grade III, Poorly Differentiated	557 (36%)	551 (36%)	
Grade Not Applicable or Not Available	129 (8.3%)	127 (8.3%)	
Cancer Stage, AJCC 6th			>0.9
Stage I	607 (39%)	584 (38%)	

Characteristic	Comparison, N = 1,551	Intervention, N = 1,537	p-value
Stage IIA	511 (33%)	515 (34%)	
Stage IIB	190 (12%)	194 (13%)	
Stage IIIA	185 (12%)	188 (12%)	
Stage IIIB	58 (3.7%)	56 (3.6%)	
WHEL Clinical Site			>0.9
Site A in California	258 (17%)	272 (18%)	
Site B in California	226 (15%)	210 (14%)	
Site C in California	267 (17%)	257 (17%)	
Site D in California	260 (17%)	256 (17%)	
Site in Arizona	242 (16%)	233 (15%)	
Site in Oregon	117 (7.5%)	116 (7.5%)	
Site in Texas	181 (12%)	193 (13%)	
Recurrence Flag			0.9
Invasive Breast Cancer Event	262 (17%)	256 (17%)	
No Invasive Breast Cancer Events	1,289 (83%)	1,281 (83%)	
Years from Diagnosis to WHEL Study Entry	1.76 (1.05, 2.81)	1.84 (1.04, 2.80)	0.8
Years from Study Entry to Recurrence/New Primary, or to Censor	7.12 (5.87, 8.44)	7.11 (5.83, 8.53)	>0.9
Years from Diagnosis to Recurrence/New Primary, or to Censor	8.97 (7.36, 10.64)	9.00 (7.35, 10.67)	0.8
Years from Diagnosis to Death or Censor	9.18 (7.76, 10.87)	9.28 (7.82, 10.89)	0.6

Table 5 Intervention Effects on All-Cause Mortality by Baseline Demographic and Clinical Characteristics

Preprocessing

```

endpoint.data = read_excel(paste0(path, 'endpoints.xls'))
year4.data = read_excel(paste0(path, 'healthstaty4.XLS'))

endpoint.data$recur_flag = as.factor(endpoint.data$recur_flag)

demo = read_excel(paste0(path, 'demographics.xls'))
phbase = read_excel(paste0(path, 'phbase.xls'))
nds = read_excel(paste0(path, "ndsfoody4.xls"))
medical = read_excel(paste0(path, "Medical.xls"))

# We need the following variables:

## Survival time
SurvTime = as.numeric(endpoint$`Years from Diagnosis to Death or Censor`)
## Group and Status
Group = as_factor(endpoint$`Intervention Group`)
Group = relevel(Group, ref = "Comparison")
a = as_factor(endpoint$`Vitality Status as of 6/1/2006`)
Status = ifelse(a == "Alive", 0, 1)
Status = as.factor(Status)
## Age at randomization, y

```

```

AgeIdx = ifelse(demo$`age at rand` < 55, "<55", ">=55") # Age indicator (<=55 or not)
## Cancer stage at diagnosis
a = endpoint$`Cancer Stage, AJCC 6th`
CancerStage = as_factor(a)
## Hormone receptor status
a = medical$`Estr Recep`
b = medical$`Prog Recep`
HormoneRecep = ifelse(a==1 & b==1, "ER+/PR+",
                      ifelse(a==1 & b==0, "ER+/PR-",
                              ifelse(a==0 & b==1, "ER-/PR+",
                                      ifelse(a==0 & b==0, "ER-/PR-", NA))))

## Time from Diag to Rand
a = endpoint$`Years from Diagnosis to WHEL Study Entry`
TimeDiagRand = as.numeric(ifelse(a <=1, 0,
                                  ifelse(a <=2, 1,
                                          ifelse(a <=3, 2, 3
                                                  )))) # Time from diagnosis to randomization

## Tumor differentiation
a = endpoint$`Cancer Grade`
TumorDiff = as_factor(a)
## No. of positive nodes (Number axillary lymph nodes positive)
a = medical$`Node Pos`
PosNodes = ifelse(a==0, 0,
                  ifelse(a < 3, 1,
                          ifelse(a < 6, 2, 3)))

## Tumor size
a = medical$`Tumor Size`
TumorSize = ifelse(a < 2, 0,
                  ifelse(a < 3, 1,
                          ifelse(a < 4, 2,
                                  ifelse(a < 5, 3, 4))))

## Physical activity
a = phbase$`NEW METS`
PhysicalAct = ifelse(a <= 210, "<210",
                    ifelse(a <= 615, "211~615",
                            ifelse(a <= 1290, "616~1290", ">1290")))

## Energy intake
b = matrix(NA, nrow = length(a), ncol=1)
colnames(b) = "KCal"
b[endpoint$ID %in% nds$ID] = nds$Kcal
KCal = as_factor(ifelse(b <= 1430, "<1430",
                        ifelse(b <= 1680, "1430~1680",
                                ifelse(b <= 1980, "1681~1980",
                                        ifelse(b > 1980, ">1980", NA)))))

##### PUT THEM TOGETHER #####
AllCauseMortalityData = data.frame(
  SurvTime, Group, Status, AgeIdx, CancerStage, HormoneRecep, TimeDiagRand,
  TumorDiff, PosNodes, TumorSize, PhysicalAct, KCal
)
#####

```

```
AllCauseMortalityData %>%
  tbl_summary(by = 'Group',
              missing_text = "Missing") %>%
  add_p()
```

```
## Table printed with 'knitr::kable()', not {gt}. Learn why at
## http://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include 'message = FALSE' in code chunk header.
```

Characteristic	Comparison, N = 1,551	Intervention, N = 1,537	p-value
SurvTime	9.18 (7.76, 10.87)	9.28 (7.82, 10.89)	0.6
Status			0.9
0	1,391 (90%)	1,381 (90%)	
1	160 (10%)	156 (10%)	
AgeIdx			>0.9
<55	917 (59%)	908 (59%)	
>=55	634 (41%)	629 (41%)	
CancerStage			>0.9
Stage IIA	511 (33%)	515 (34%)	
Stage I	607 (39%)	584 (38%)	
Stage IIB	190 (12%)	194 (13%)	
Stage IIIA	185 (12%)	188 (12%)	
Stage IIIB	58 (3.7%)	56 (3.6%)	
HormoneRecep			0.054
ER-/PR-	319 (21%)	299 (20%)	
ER-/PR+	77 (5.1%)	52 (3.5%)	
ER+/PR-	170 (11%)	198 (13%)	
ER+/PR+	950 (63%)	958 (64%)	
Missing	35	30	
TimeDiagRand			>0.9
0	350 (23%)	352 (23%)	
1	508 (33%)	488 (32%)	
2	372 (24%)	375 (24%)	
3	321 (21%)	322 (21%)	
TumorDiff			>0.9
Grade Not Applicable or Not Available	129 (8.3%)	127 (8.3%)	
Grade II, Moderately Differentiated	620 (40%)	620 (40%)	
Grade I, Well Differentiated	245 (16%)	239 (16%)	
Grade III, Poorly Differentiated	557 (36%)	551 (36%)	
PosNodes			0.7
0	896 (58%)	879 (57%)	
1	374 (24%)	357 (23%)	
2	164 (11%)	169 (11%)	
3	117 (7.5%)	131 (8.5%)	
Missing	0	1	
TumorSize			0.8
0	769 (50%)	752 (49%)	
1	441 (28%)	421 (27%)	
2	160 (10%)	174 (11%)	
3	72 (4.7%)	79 (5.1%)	
4	106 (6.8%)	109 (7.1%)	

Characteristic	Comparison, N = 1,551	Intervention, N = 1,537	p-value
Missing	3	2	0.3
PhysicalAct			
<210	407 (26%)	443 (29%)	
>1290	387 (25%)	351 (23%)	
211~615	383 (25%)	368 (24%)	
616~1290	374 (24%)	375 (24%)	0.6
KCal			
<1430	454 (38%)	446 (40%)	
1430~1680	308 (26%)	287 (25%)	
>1980	174 (14%)	144 (13%)	
1681~1980	271 (22%)	249 (22%)	
Missing	344	411	