What Caused the Decline in BP Control After 2013?

Analysis Based on NHANES Survey Data

Xianjie Que

Study Background & Objective

- Data: NHANES 1999–2020 (10 biennial cycles)
- Begin_year: 1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013, 2015, 2017
- BP control rate declined after 2013 (previously rising trend)

- **Objective**: Understand the decline in BP control since 2013.

Raw Dataset

- raw dataset: 26,757 obs, 160 variables
- Response variable: "*bp_control_accaha*"

(binary: 1 = control, 0 = no control)

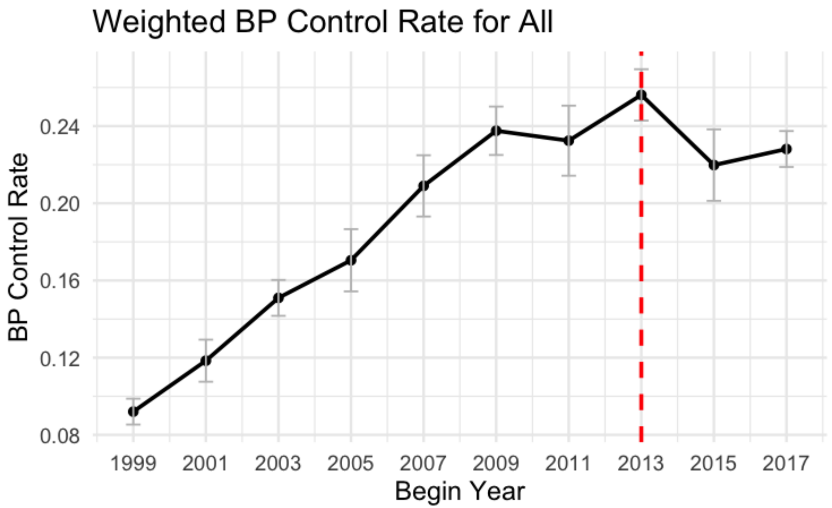
cardioStatsUSA 0.0.1 RESULTS GET STARTED TECHNICAL STUFF CHANGELOG

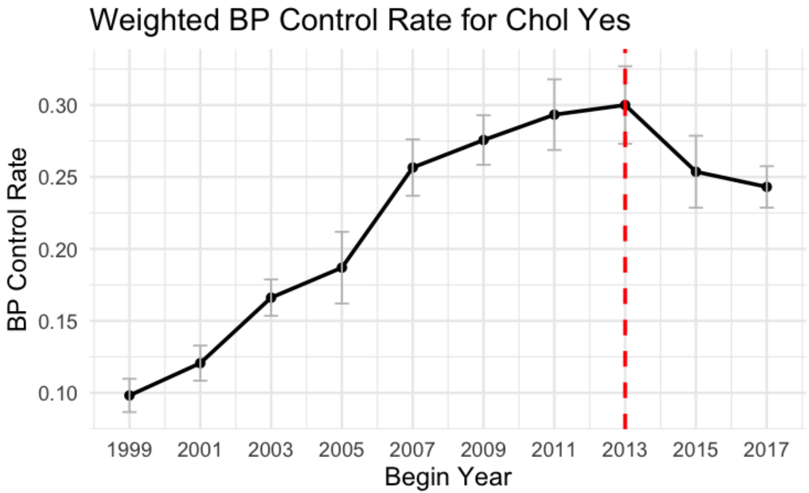
bp_control_accaha

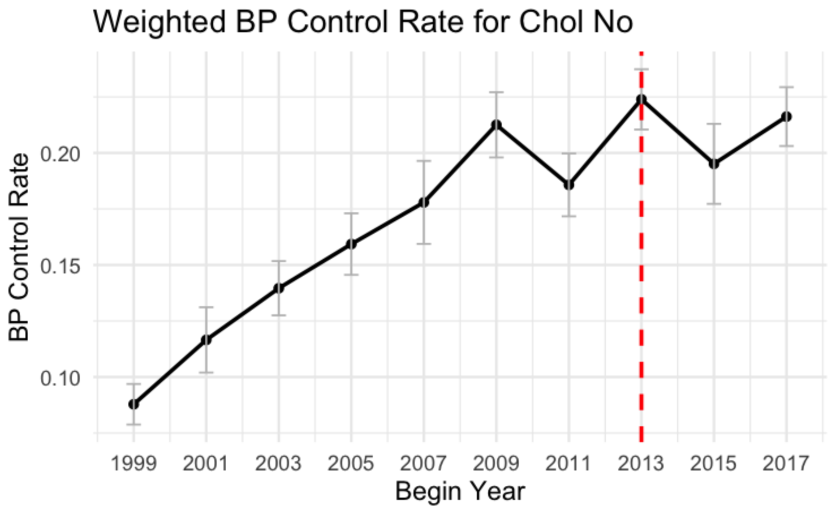
- **Label**: Blood pressure control defined by the 2017 ACC/AHA BP guideline
- **Description**: Systolic and diastolic blood pressure controlled to the levels recommended in the 2017 ACC/AHA BP guideline, systolic blood pressure < 130 mm Hg and diastolic blood pressure < 80 mm Hg except for those >= 65 years of age without diabetes, chronic kidney disease, history of cardiovascular disease or 10-year predicted ASCVD risk >= 10% estimated using the Pooled Cohort risk equations. For this group, blood pressure control was defined as systolic blood pressure < 130 mm Hg

Analysis Steps

1. Add phase variable (Before / After 2013)
2. Split into two subsets (With / Without cholesterol data)
3. Data cleaning
4. Imputation
5. LASSO with interaction terms
6. Weighted logistic regression with interactions
7. Trend plots



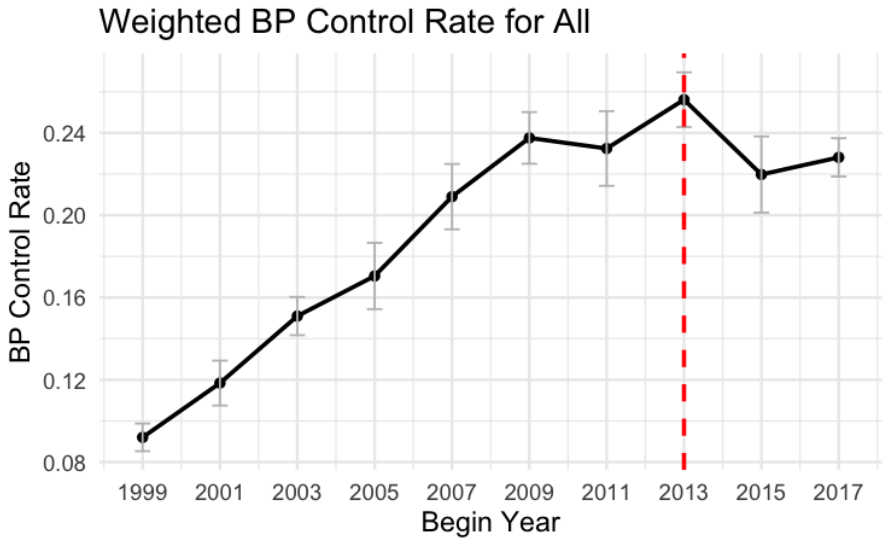




Decline after 2013 is clearer in the cholesterol subpopulation

1. Add Phase Variable

- Define phase based on clear national trends in BP control:
  - "*phase"* = Rising (before 2013)
  - "*phase"* = Falling (after 2013)



2. Split into Two Subsets

- NHANES variable "*svy_subpop_chol*" divides sub-population (26,757 total):
  - With cholesterol data ("*svy_subpop_chol*" == 1) → 11,118 people
  - Without cholesterol data ("*svy_subpop_chol*" == 0) → 15,639 people (58%)

Currently, two sub-populations are included in `nhanes_data`:

- blood pressure and hypertension (n = 56,017)
- lipids and cholesterol (n = 25,151)

To access a sub-population, simply filter `nhanes_data` to contain rows where the corresponding sub-population indicator is equal to 1, e.g.,

```
# for blood pressure and hypertension sub-population
nhanes_data[svy_subpop_htn == 1]

# for lipids and cholesterol sub-population
nhanes_data[svy_subpop_chol == 1]
```

If I use full dataset:

- 41 Cholesterol-related variables with ~58% missing rate
- Have to remove these variables

```r
### check "chol_" variables

```{r check "chol_" variables}
chol_vars <- grep("^chol_", names(prep), value = TRUE)
length(chol_vars) # 41

missing_rates <- sapply(prep[chol_vars], function(col) {
  mean(is.na(col))
})

range(missing_rates) # 0.5777927 0.5934896

# transform to a data.frame and order by missing rate (descending)
missing_df <- data.frame(
```

137:23 Chunk 5: check "chol_" variables

Console | Terminal | Background Jobs

R 4.4.1 · ~/Desktop/1/SP2025.660 Biomedical Data Mining/final project/

```
> length(chol_vars) # 41
[1] 41
>
> missing_rates <- sapply(prep[chol_vars], function(col) {
+   mean(is.na(col))
+ })
>
> range(missing_rates) # 0.5777927 0.5934896
[1] 0.5777927 0.5934896
```

missing_df

Filter

	Variable	MissingRate
chol_measured_last	chol_measured_last	0.5934896
chol_trig	chol_trig	0.5928542
chol_nonhdl	chol_nonhdl	0.5901259
chol_hdl	chol_hdl	0.5877341
chol_measured_never	chol_measured_never	0.5865381
chol_ldl_5cat	chol_ldl_5cat	0.5846694
chol_ldl	chol_ldl	0.5839967
chol_total	chol_total	0.5839593
chol_nonhdl_5cat	chol_nonhdl_5cat	0.5836977
chol_med_recommended_ever	chol_med_recommended_ever	0.5784281
chol_med_use_sr	chol_med_use_sr	0.5782786
chol_total_gteq_200	chol_total_gteq_200	0.5777927
chol_total_gteq_240	chol_total_gteq_240	0.5777927
chol_hdl_low	chol_hdl_low	0.5777927
chol_trig_gteq_150	chol_trig_gteq_150	0.5777927
chol_ldl_lt_70	chol_ldl_lt_70	0.5777927
chol_ldl_gteq_70	chol_ldl_gteq_70	0.5777927
chol_ldl_lt_100	chol_ldl_lt_100	0.5777927
chol_ldl_gteq_100	chol_ldl_gteq_100	0.5777927
chol_ldl_gteq_190	chol_ldl_gteq_190	0.5777927
chol_ldl_persistent	chol_ldl_persistent	0.5777927
chol_nonhdl_lt_100	chol_nonhdl_lt_100	0.5777927

Showing 1 to 23 of 41 entries, 2 total columns
```

2. Split into Two Subsets

Cholesterol
HDL
Triglyceride
LDL
Total Chol

But cholesterol is critical:

- Cholesterol plays a key role in lipid metabolism
- Dyslipidemia co-occurs with hypertension (60–64% in NHANES)
- Metabolic syndrome (abdominal obesity, ↑triglycerides, ↓HDL) linked to uncontrolled BP
- **Key markers**: high triglycerides, Low HDL (linked to uncontrolled hypertension)

3. Data Cleaning

3.1 Fake Missing Values

Convert blank/"NA" strings to real `NA`.

```
Numeric   cleaned: svy_id              | NA: 0 → 3
Numeric   cleaned: svy_strata          | NA: 0 → 964
Numeric   cleaned: demo_age_years      | NA: 0 → 359
Numeric   cleaned: bp_sys_mean         | NA: 0 → 4
Numeric   cleaned: bp_dia_mean         | NA: 0 → 368
Character cleaned: chol_measured_never  | NA: 0 → 15694
Character cleaned: chol_measured_last   | NA: 0 → 15880
Numeric   cleaned: chol_total          | NA: 15618 → 15625
Character cleaned: chol_total_gteq_200  | NA: 0 → 15460
Character cleaned: chol_total_gteq_240  | NA: 0 → 15460
Numeric   cleaned: chol_hdl            | NA: 15617 → 15726
Character cleaned: chol_hdl_low         | NA: 0 → 15460
Numeric   cleaned: chol_trig           | NA: 15626 → 15863
Character cleaned: chol_trig_gteq_150   | NA: 0 → 15460
Character cleaned: chol_ldl_5cat        | NA: 0 → 15644
Character cleaned: chol_ldl_lt_70       | NA: 0 → 15460
Character cleaned: chol_ldl_gteq_70     | NA: 0 → 15460
Character cleaned: chol_ldl_lt_100      | NA: 0 → 15460
Character cleaned: chol_ldl_gteq_100    | NA: 0 → 15460
Character cleaned: chol_ldl_gteq_190    | NA: 0 → 15460
Character cleaned: chol_ldl_persistent  | NA: 0 → 15460
Numeric   cleaned: chol_nonhdl         | NA: 15618 → 15790
Character cleaned: chol_nonhdl_5cat     | NA: 0 → 15618
Character cleaned: chol_nonhdl_lt_100   | NA: 0 → 15460
Character cleaned: chol_nonhdl_gteq_100 | NA: 0 → 15460
Character cleaned: chol_nonhdl_gteq_220 | NA: 0 → 15460
Character cleaned: chol_med_use         | NA: 0 → 15460
Character cleaned: chol_med_use_sr      | NA: 0 → 15473
Character cleaned: chol_med_statin      | NA: 0 → 15460
Character cleaned: chol_med_ezetimibe   | NA: 0 → 15460
Character cleaned: chol_med_pcsk9i      | NA: 0 → 15460
Character cleaned: chol_med_bile        | NA: 0 → 15460
Character cleaned: chol_med_fibric_acid | NA: 0 → 15460
Character cleaned: chol_med_atorvastatin | NA: 0 → 15460
Character cleaned: chol_med_simvastatin | NA: 0 → 15460
Character cleaned: chol_med_rosuvastatin | NA: 0 → 15460
```

3. Data Cleaning

3.2 Variable Removal Criteria

3.2.1 High collinearity ≥ 0.95 —— first Collinearity Check

3.2.2 Unit conversions for lab variables (e.g. SI vs traditional units)

3.2.3 Drop variables that do not vary or barely change

3.2.4 Remove actual BP measures

3.2.5 Drop survey design variables only for LASSO, re-add for logistic

3. Data Cleaning

3.2.1 High collinearity ≥ 0.95 —— first Collinearity Check

Keep variables with lower missing rate and clinical meaning

var1 <chr>	var2 <chr>	r <dbl>
LBDHDL	chol_hdl	1.0000000
LBDSTRSI	chol_trig	0.9822510
LBDLDL	chol_ldl	0.9991425
FriedewaldLDL	chol_ldl	0.9975579
ldl_corrected	chol_ldl	1.0000000
LBXGH	cc_hba1c	1.0000000
Survey_Year	YEAR	1.0000000
SDDSRVYR	YEAR	1.0000000
SDMVSTRA	YEAR	0.9943459
Begin_Year	YEAR	1.0000000

21–30 of 42 rows   Previous 1 2 3 4 5 Next

3. Data Cleaning

3.2.1 High collinearity ≥ 0.95 —— first Collinearity Check

Key Cholesterol Variables

Marker Type	Keep	Remove
HDL	chol_hdl	LBDHDL
Triglyceride	chol_trig	LBDSTRSI
LDL	ldl_corrected	LBDLDL, chol_ldl
Total Chol	LBXTC	LBXSCH, chol_total

3. Data Cleaning

3.2.2 Unit conversions for lab variables (SI vs traditional units)

Keep	SI unit (e.g., LBDSTPSI)	mol/L mmol/L
Remove	traditional unit (e.g., LBXSTP)	mg/dL

Description: df [5 x 3]

LBX_variable <chr>	LBD_variable <chr>	correlation <dbl>
LBXSTP	LBDSTPSI	1
LBXSTR_x	LBDSTRSI	1
LBXSUA	LBDSUASI	1
LBXSGB	LBDSGBSI	1
LBXSTR_y	LBDSTRSI	1

3. Data Cleaning

```
[1] "svy_subpop_chol" "htn_accaha"       "RIDSTATR"
svy_subpop_chol        htn_accaha          RIDSTATR
            "1"             "Yes"               "2"
```

3.2.3 Drop variables that:

- Do not vary (only one value)
- Barely change (>99% same value)

```
>
> print(low_variability_vars)
[1] "chol_med_pcsk9i"        "chol_med_bile"
[3] "chol_med_pitavastatin" "chol_med_fluvastatin"
[5] "chol_med_other"         "BPAARM"
>
> # View the dominant values and proportions of these variables
> sapply(prep_chol[low_variability_vars], function(x) {
+   non_missing <- na.omit(x)
+   top_val <- names(sort(table(non_missing), decreasing = TRUE))[1]
+   top_prop <- max(table(non_missing)) / length(non_missing)
+   paste0("Most frequent value: ", top_val, " (", round(top_prop * 100, 2), "%)")
+ })
                   chol_med_pcsk9i
"Most frequent value: No (99.96%)"
                     chol_med_bile
"Most frequent value: No (99.68%)"
             chol_med_pitavastatin
"Most frequent value: No (99.95%)"
              chol_med_fluvastatin
"Most frequent value: No (99.75%)"
                    chol_med_other
"Most frequent value: No (99.42%)"
                            BPAARM
 "Most frequent value: 1 (99.22%)"
>
```

3. Data Cleaning

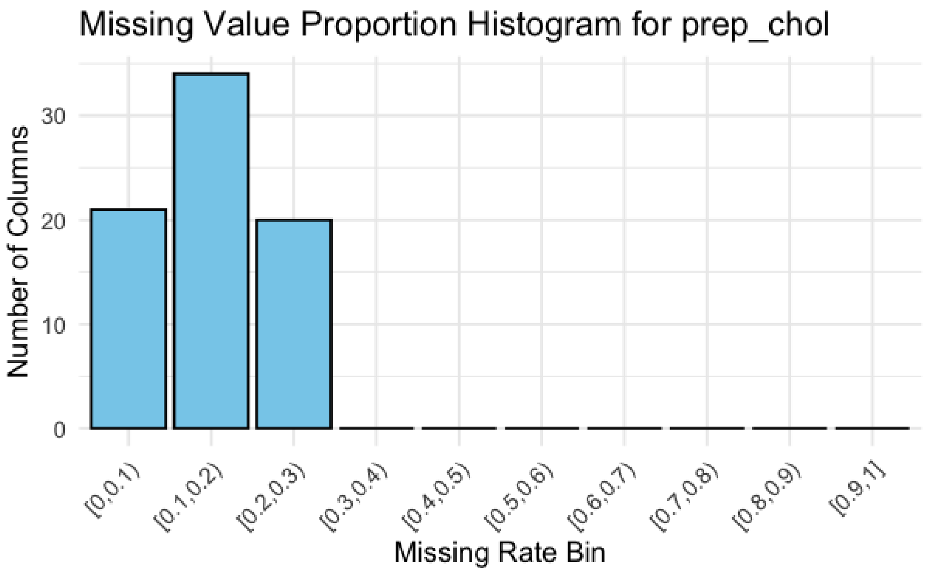
3.2.4 Remove actual BP measures to avoid tautological predictions

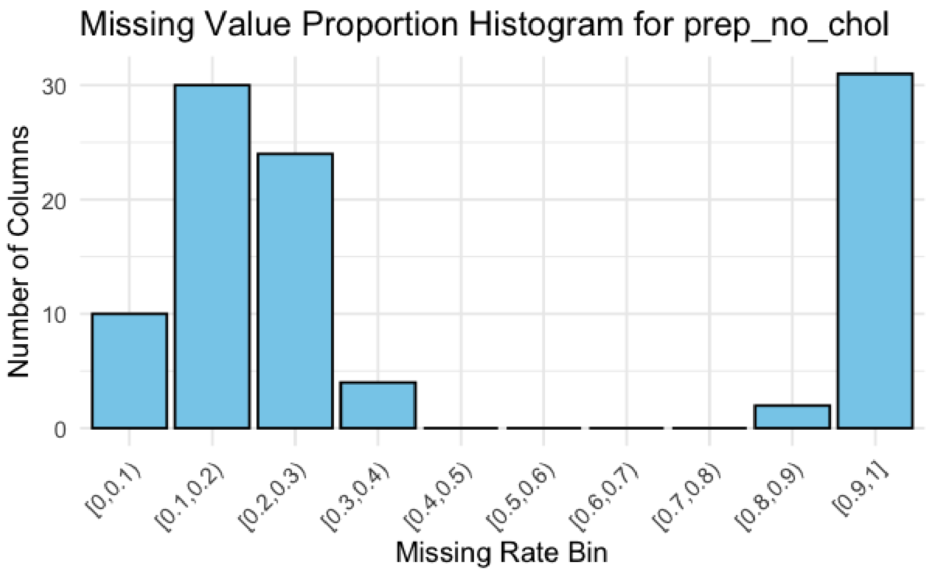
- Systolic blood pressure: BPXSY1, BPXSY2, BPXSY3, bp_sys_mean
- Diastolic blood pressure: BPXDI1, BPXDI2, BPXDI3, bp_dia_mean

3.2.5 Drop survey design variables only for LASSO, re-add for logistic regression

4. Imputation

1. Visualize missing rate distribution
2. Threshold = 0.5: remove columns with >50% missing
3. Impute:
   - Continuous variables: impute using EM algorithm (m = 1)
   - Categorical variables: impute using KNN (k = 5)





Second Collinearity Check After Imputation

- Before imputation: r>=0.95 (ensure numerical stability)
- After imputation: r>=0.8 (remove redundancy)

BMI Group

Keep	Remove
BMXBMI	BMXWT
BMXWAIST	BMXARMC
	BPACSZ

Glucose vs HbA1c

Keep	Remove
cc_hba1c	LBXSGL

5. LASSO with Interaction Terms

- Dataset: 11,118 rows × 108 columns
- Formula: main effects + interactions

  bp_control_accaha ~ x1 + x2 + x3 + x1:phase + x2:phase + x3:phase + ...

- Reference: $phase = Rising$
- Interaction terms "*x1:phase*" represent phase-based differences in effects

5. LASSO with Interaction Terms

```
interaction terms from Lasso:
 [1] "chol_trig:phaseFalling"
 [2] "LBDSPH:phaseFalling"
 [3] "LBDSCR:phaseFalling"
 [4] "cc_egfr_lt60Yes:phaseFalling"
 [5] "cc_acr_gteq30Yes:phaseFalling"
 [6] "demo_age_cat75+:phaseFalling"
 [7] "demo_raceNon-Hispanic Black:phaseFalling"
 [8] "demo_raceNon-Hispanic White:phaseFalling"
 [9] "htn_resistant_jnc7Yes:phaseFalling"
[10] "htn_resistant_jnc7_thzYes:phaseFalling"
[11] "chol_nonhdl_5cat100 to <130 mg/dL:phaseFalling"
[12] "chol_med_rosuvastatinYes:phaseFalling"
[13] "chol_med_addon_useYes:phaseFalling"
[14] "cc_diabetesYes:phaseFalling"
```

5. LASSO with Interaction Terms

- Metabolism and renal function: chol_trig, LBDSPH, cc_egf_lt60, cc_diabetes……
- Lipid-lowering therapy: chol_nonhdl_5cat(100 to <130 mg/dL), chol_med_rosuvastatin, chol_med_addon_use
- Demographic variables: demo_age_cat(75+), demo_race(Non_Hispanic_Black), demo_race(Non_Hispanic_White),
- hypertension resistance and medication: htn_resistant_jnc7, htn_resistant_jnc7_thz

6. Weighted logistic regression with interactions

- Use weighted logistic regression ($svyglm$) on selected interaction terms.
- Formula: only selected interactions

  bp_control_accaha ~ x1:phase + x2:phase + x3:phase + ...

NHANES Survey Design

cardioStatsUSA 0.0.1 RESULTS GET STARTED TECHNICAL STUFF CHANGELOG

Survey variables

In each NHANES cycle, potential participants were identified using a multi-stage sampling process. The variables below are based on this process.

svy_id

- **Label**: Participant identifier
- **Description**: NHANES participant unique identifier.

svy_psu

- **Label**: Primary sampling unit
- **Description**: Population sampling unit. This variable is used to account for the non-random selection of study participants for NHANES

svy_strata

- **Label**: Strata
- **Description**: Population stratification. This variable is used to account for the non-random selection of study participants for NHANES

svy_weight_mec

- **Label**: Mobile examination center weights
- **Description**: Weight applied to produce statistical estimates for the non-institutionalized US population. This weight is used for calculating means and proportions.

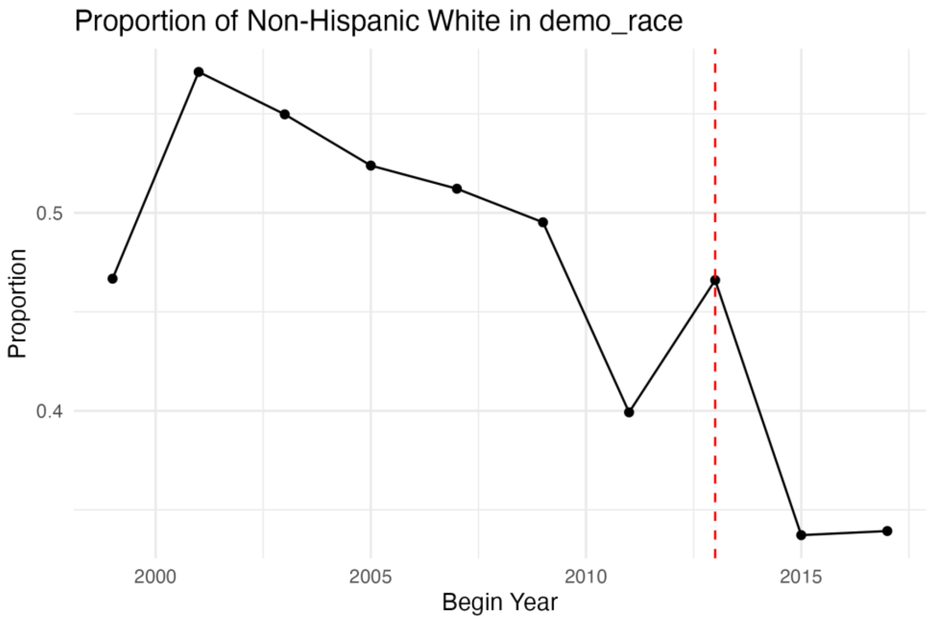
```
design <- svydesign(ids = ~svy_psu,
                    strata = ~svy_strata,
                    weights = ~svy_weight_mec,
                    nest = TRUE,
                    data = prep_svy)

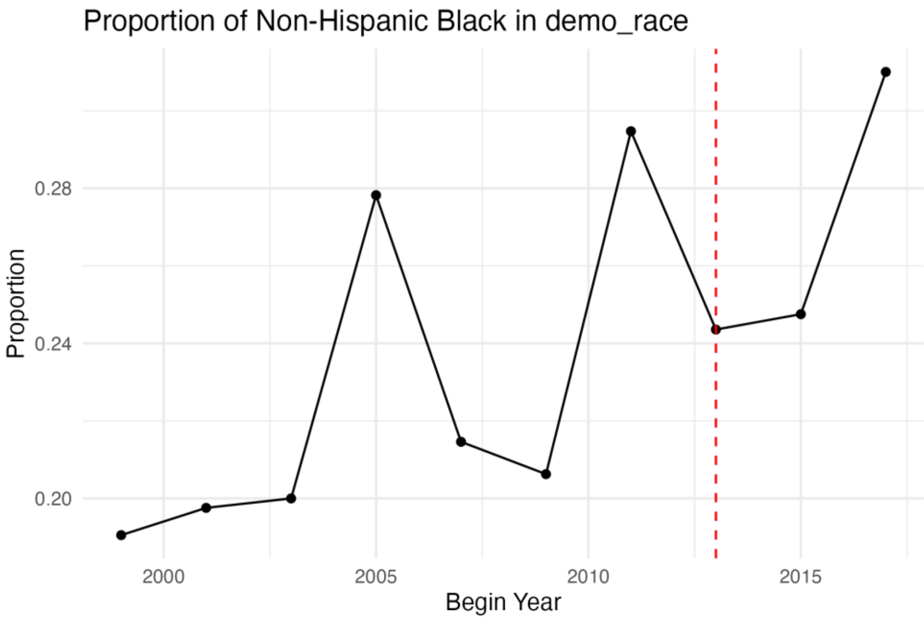
fit_svy <- svyglm(formula_obj,
                  design = design,
                  family = binomial())
```

7. Trend Plots

- Plot trends of selected variables by year
- Focus on patterns before vs. after 2013

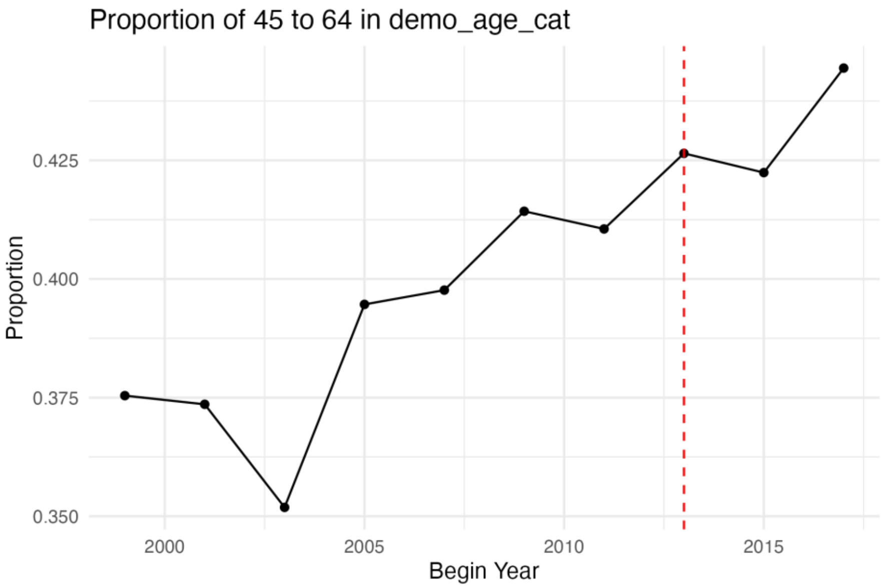
The proportion of White individuals declined significantly after 2013, while the proportion of Black individuals increased.

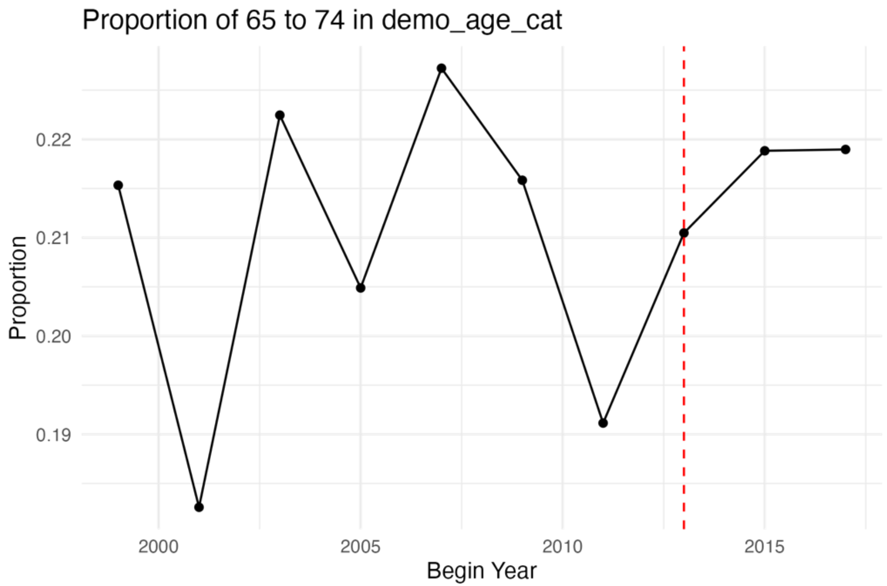




Salt-sensitive hypertension is more common among African Americans, and it is generally more difficult to control.

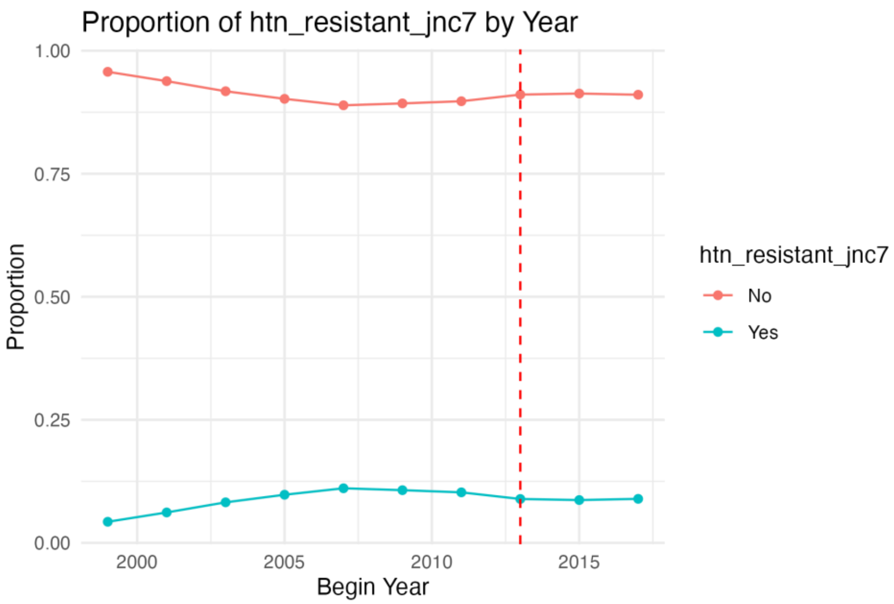
Not a drastic change, but the proportion of people aged 45-74 has increased after 2013.

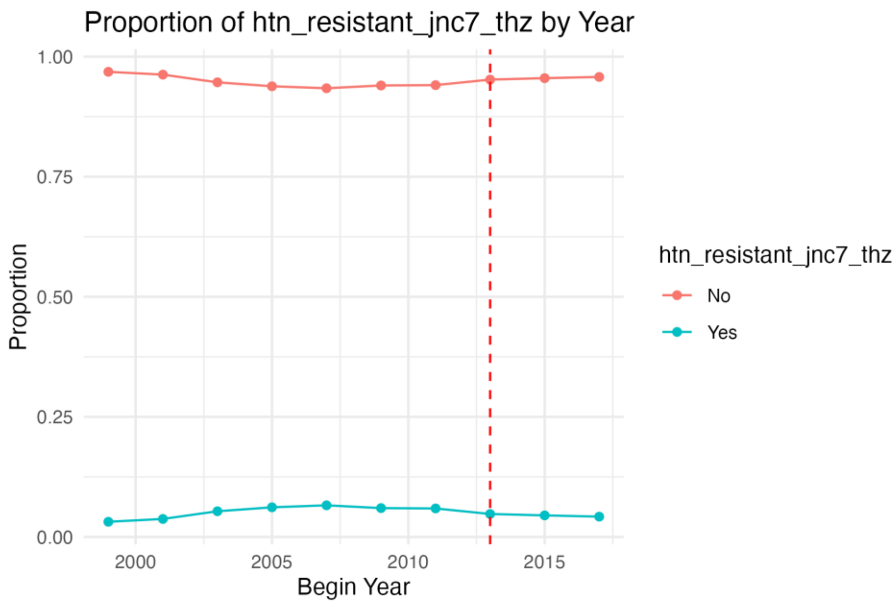




In contrast, the proportion of people aged 75 and above and 18-45 years old has decreased.

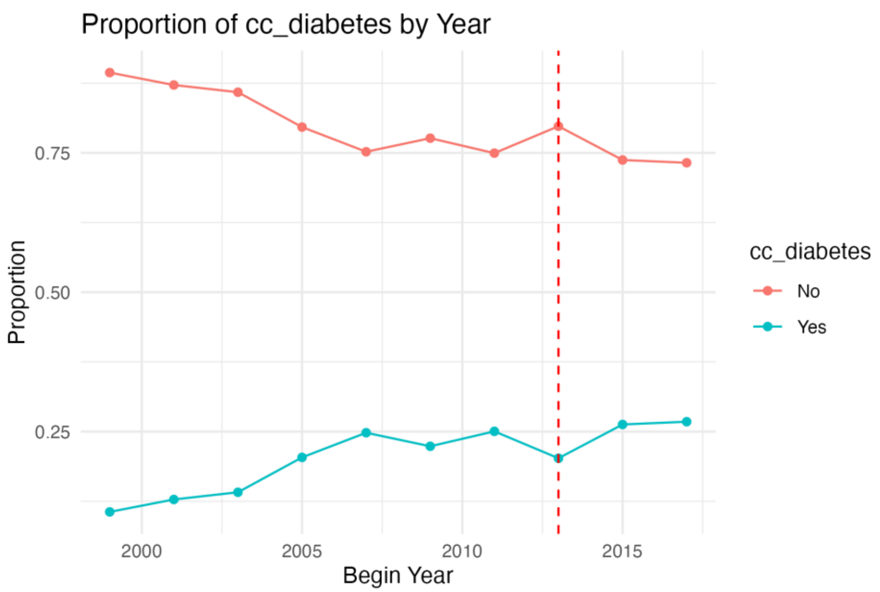
Interestingly, variables related to hypertension resistance and medication use did not change significantly, suggesting they contributed little to the decline in BP control.

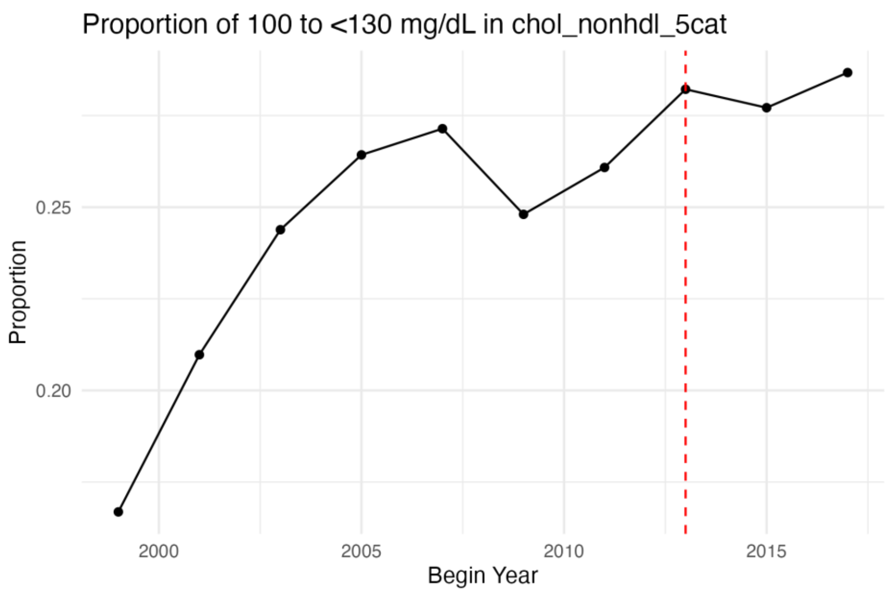




Maybe they are more cautious about managing their blood pressure.

For metabolism-related variables, the proportion of individuals with diabetes increased beginning in 2013.





Repeat Process on Subset Without Cholesterol

- Apply same steps: data cleaning, imputation, Lasso, Logistic Regression, Trend Plots
- Compare variables selected across subsets.

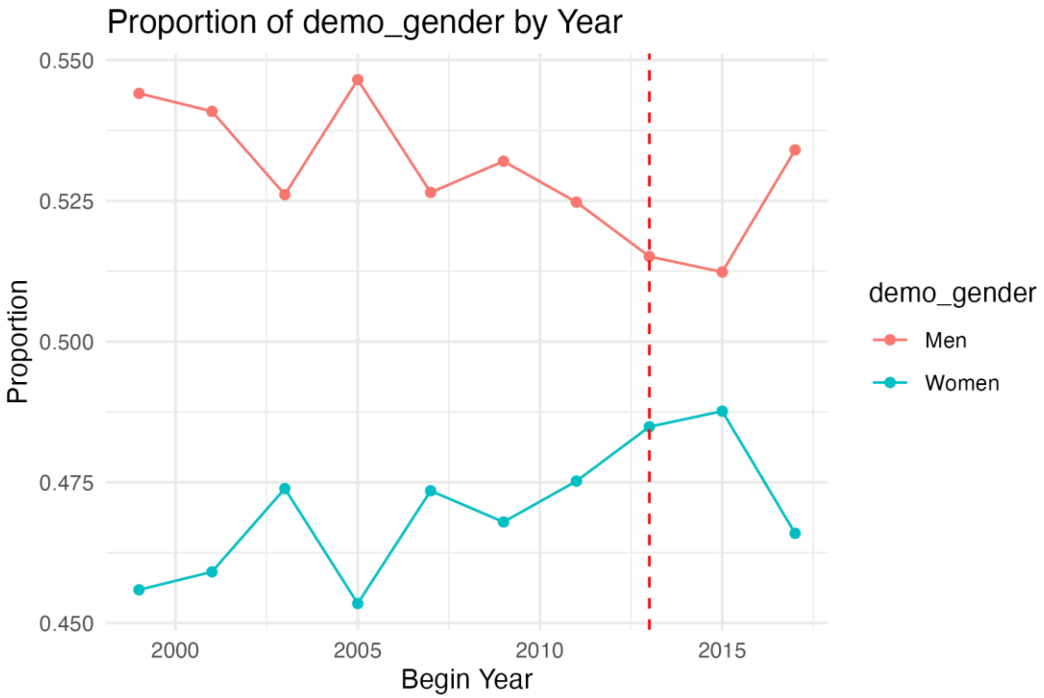
5. LASSO with Interaction Terms

```
interaction terms from Lasso:
 [1] "cc_acr:phaseFalling"
 [2] "URXUMA:phaseFalling"
 [3] "LBXSGTSI:phaseFalling"
 [4] "LBXSIR:phaseFalling"
 [5] "LBXPLTSI:phaseFalling"
 [6] "LBDSCR:phaseFalling"
 [7] "cc_cvd_miYes:phaseFalling"
 [8] "cc_cvd_chdYes:phaseFalling"
 [9] "cc_cvd_hfYes:phaseFalling"
[10] "demo_age_cat65 to 74:phaseFalling"
[11] "demo_raceNon-Hispanic Asian:phaseFalling"
[12] "demo_raceNon-Hispanic Black:phaseFalling"
[13] "demo_raceNon-Hispanic White:phaseFalling"
[14] "demo_genderWomen:phaseFalling"
[15] "htn_resistant_accaha_thzYes:phaseFalling"
[16] "cc_diabetesYes:phaseFalling"
[17] "cc_ckdYes:phaseFalling"
[18] "cc_cvd_anyYes:phaseFalling"
```

5. LASSO with Interaction Terms

- Demographic variables: demo_age_cat(65 to 74), demo_race(all races), demo_gender(Women)
- hypertension resistance and medication: htn_resistant_jnc7_thz
- Metabolism and renal function: cc_diabetes, cc_ckd, cc_acr, LBDSCR, ……
- Cardiovascular disease-related variables: cc_cvd_mi, cc_cvd_chd, cc_cvd_hf, cc_cvd_any

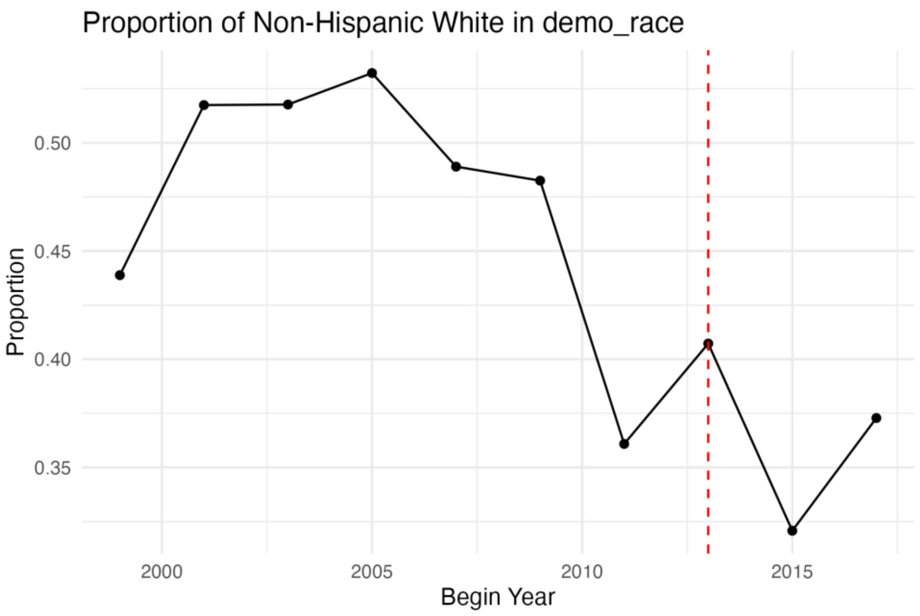
After 2015, the proportion of female declined

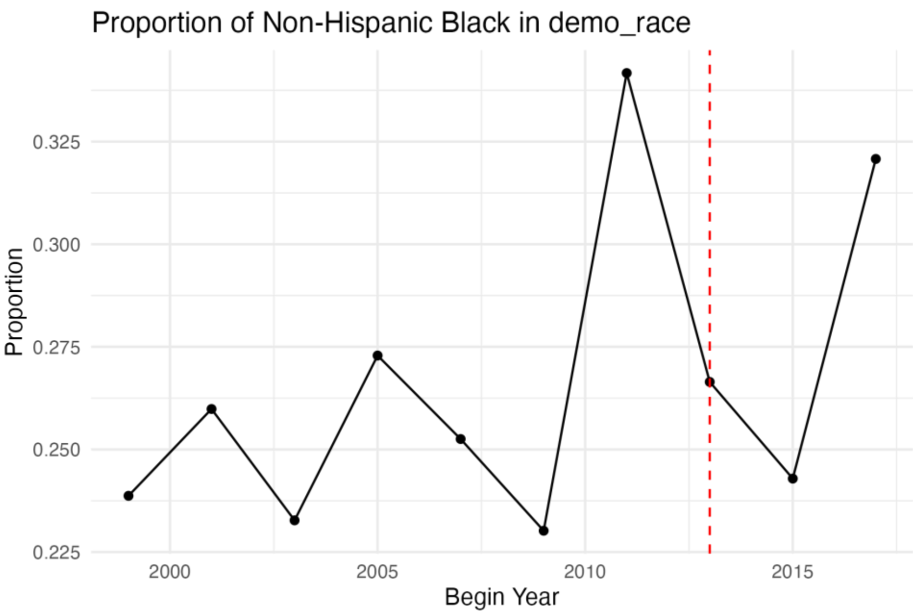


There is evidence that men are more frequently reported to have poor blood pressure control, especially in middle-aged and older populations.

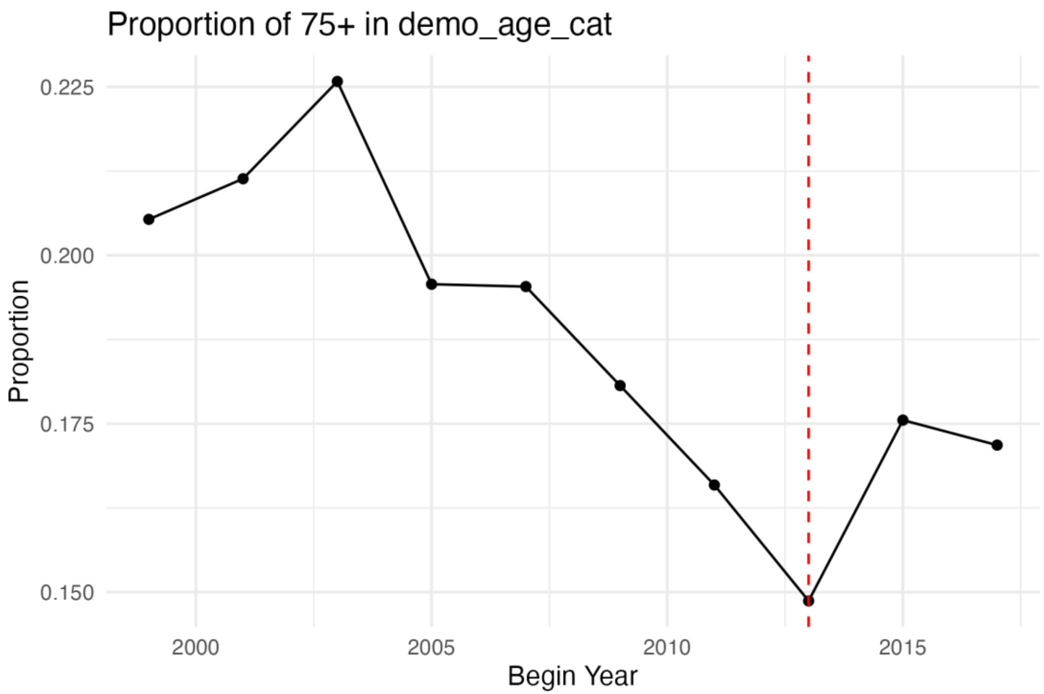
Everett B, Zajacova A. Gender differences in hypertension and hypertension awareness among young adults. Biodemography Soc Biol. 2015;61(1):1-17. doi: 10.1080/19485565.2014.929488. PMID: 25879259; PMCID: PMC4896734.

Similarly, the proportion of White individuals significantly declined after 2013, while the proportion of Black individuals increased markedly after 2015.

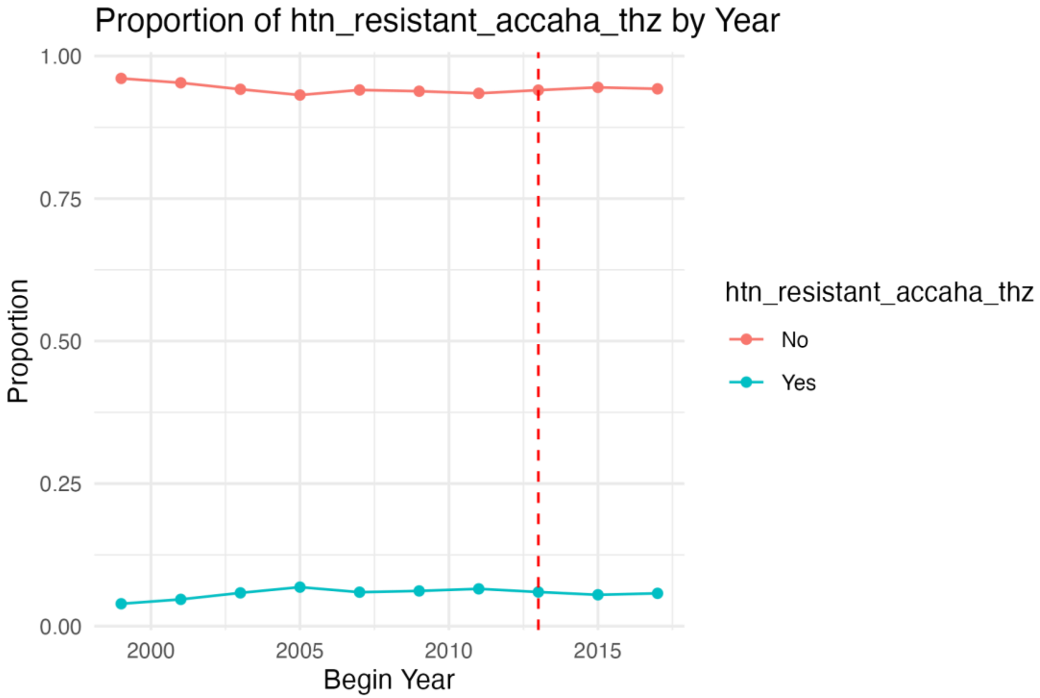




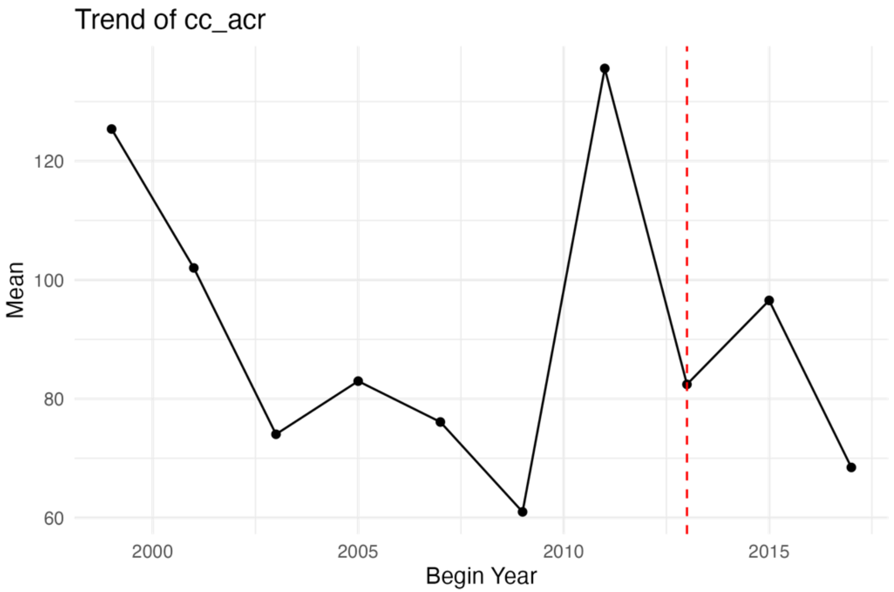
The proportion of people aged 75+ has increased since 2013, although not significantly.

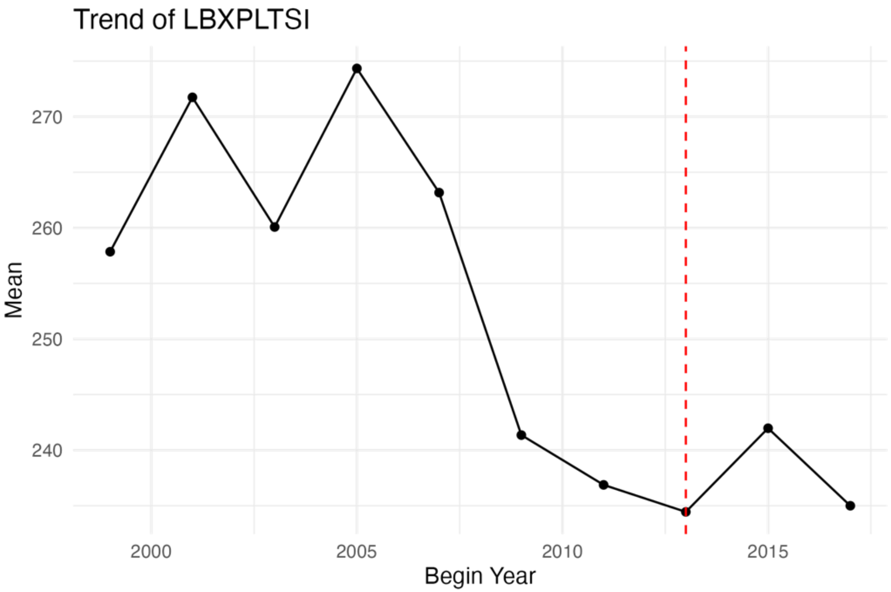


Likewise, there were no significant changes in variables related to hypertension resistance and medication.

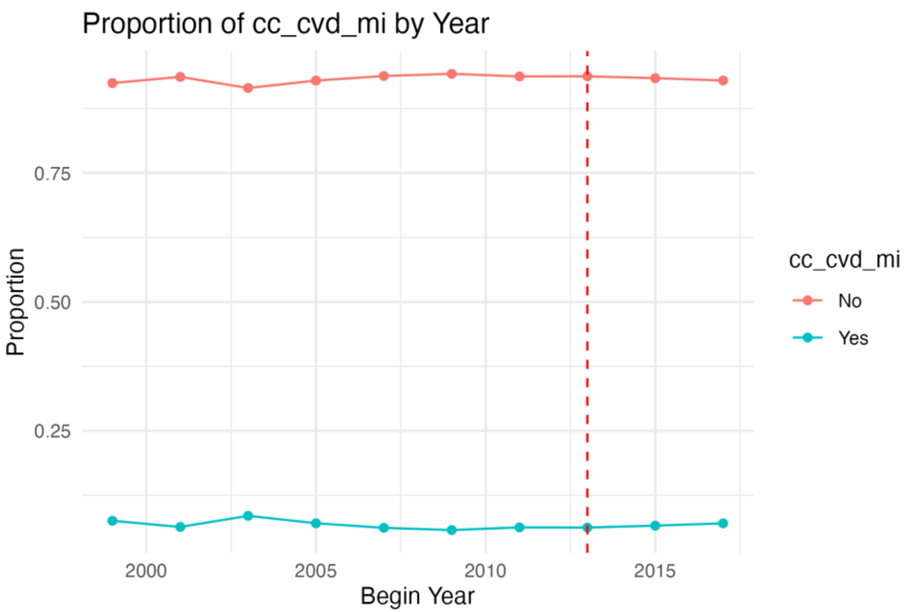


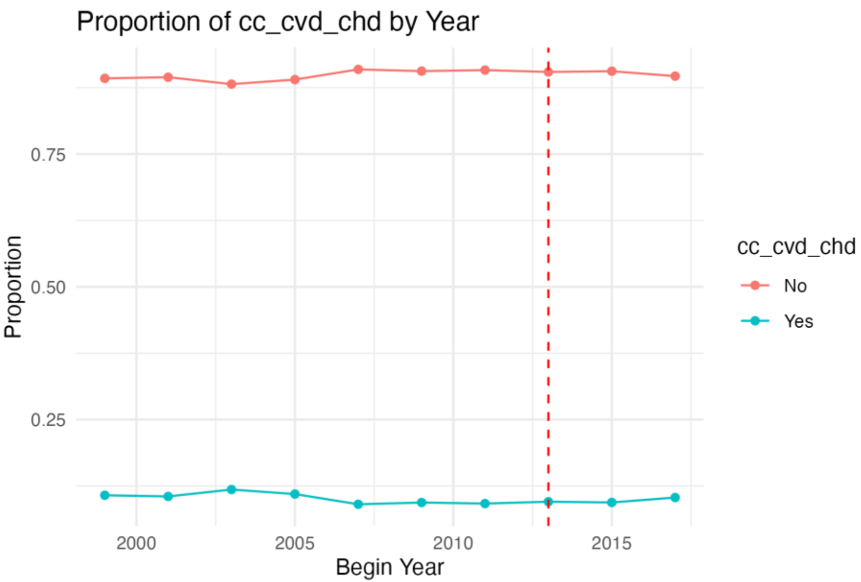
Variables associated with kidney damage and metabolic disorders have increased since 2013.





Variables related to cardiovascular disease have shown no significant changes.





Conclusion

- The decline in BP control may be associated with:
  1. Demographically, the decline in the proportion of females and White individuals, along with the increase in Black individuals
  2. a growing population at high metabolic risk. (diabetes↑, HDL↓, variables linked to kidney damage or metabolic dysregulation ↑)
- However, cardiovascular disease-related variables, variables related to hypertension resistance and medication showed little change, suggesting a minimal contribution to the decline in BP control.

Further

- Explore some factors I expect were strongly correlated but not selected:
  - obesity (“cc_bmi”, “BMXBMI”, “BMXWAIST”)
  - smoking (“cc_smoke”)

- Imputation:
  - using EM algorithm (m = 5)
  - Try other imputation methods

- Add interaction between predictors

References

- Egan BM, Li J, Qanungo S, Wolfman TE. Blood pressure and cholesterol control in hypertensive hypercholesterolemic patients: national health and nutrition examination surveys 1988-2010. Circulation. 2013 Jul 2;128(1):29-41. doi: 10.1161/CIRCULATIONAHA.112.000500. PMID: 23817481; PMCID: PMC4066305.
- Zidek, W., Naditch-Brûlé, L., Perlini, S. *et al.* Blood pressure control and components of the metabolic syndrome: the GOOD survey. *Cardiovasc Diabetol* **8**, 51 (2009). https://doi.org/10.1186/1475-2840-8-51
- Everett B, Zajacova A. Gender differences in hypertension and hypertension awareness among young adults. Biodemography Soc Biol. 2015;61(1):1-17. doi: 10.1080/19485565.2014.929488. PMID: 25879259; PMCID: PMC4896734.
- Sahinoz M, Elijovich F, Ertuglu LA, Ishimwe J, Pitzer A, Saleem M, Mwesigwa N, Kleyman TR, Laffer CL, Kirabo A. Salt Sensitivity of Blood Pressure in Blacks and Women: A Role of Inflammation, Oxidative Stress, and Epithelial Na+ Channel. Antioxid Redox Signal. 2021 Dec 20;35(18):1477-1493. doi: 10.1089/ars.2021.0212. PMID: 34569287; PMCID: PMC8713266.