

Medicare ACO Spending Patterns: Replication One-Pager

Eric Walczyk

October 09, 2025

Load and Clean Data

```
library(readr)
library(dplyr)
library(purrr)
library(stringr)
library(janitor)

files <- list.files(pattern = "(2014|2015|2016|2017).*\\.csv$", full.names = TRUE)
years <- str_extract(basename(files), "(2014|2015|2016|2017)")

# 1) read + clean + add year
datasets <- map2(files, years, ~ read_csv(.x, show_col_types = FALSE) |>
  clean_names() |>
  mutate(year = as.integer(.y)) |>
  # 2) normalize the infamous CapAnn names
  rename_with(~ str_replace_all(.x, "^cap_ann_", "capann_")) |>
  rename_with(~ str_replace_all(.x, "amb_pay$", "ambpay")) ) # PB stays PB

names(datasets) <- years

# 3) choose the columns you actually use downstream
needed <- c("aco_num", "year", "sav_rate",
  "capann_inp_all", "capann_opd", "capann_snf",
  "capann_pb", "capann_ambpay", "capann_hha",
  "capann_dme", "capann_hsp",
  "qual_score") # optional

# 4) make sure each frame has all needed cols and types match
harmonize_one <- function(df) {
  miss <- setdiff(needed, names(df))
  if (length(miss)) for (nm in miss) df[[nm]] <- NA

  df |>
  mutate(
    aco_num = as.character(aco_num),
    across(any_of(c(
      "sav_rate", "capann_inp_all", "capann_opd", "capann_snf",
      "capann_pb", "capann_ambpay", "capann_hha", "capann_dme", "capann_hsp",
      "qual_score"
    )), ~ readr::parse_number(as.character(.x)))
  ) |>
  select(any_of(needed))
}

datasets2 <- map(datasets, harmonize_one)

# 5) now you can bind safely
aco_data <- bind_rows(datasets2, .id = "src_year") |>
  mutate(src_year = as.integer(src_year))
```

```
write.csv(aco_data, "aco_data.csv")
```

Model Set Up

```

### Model Set Up
expenditure_columns <- c(
  "capann_inp_all", "capann_opd", "capann_snf",
  "capann_pb", "capann_ambpay", "capann_hha",
  "capann_dme", "capann_hsp"
)

### Remove outliers in savings rate
lower_bound <- quantile(aco_data$sav_rate, 0.25, na.rm = TRUE) - 1.5 * IQR(aco_data$sav_rate, na.rm = TRUE)
upper_bound <- quantile(aco_data$sav_rate, 0.75, na.rm = TRUE) + 1.5 * IQR(aco_data$sav_rate, na.rm = TRUE)

aco_data <- aco_data %>%
  filter(sav_rate >= lower_bound, sav_rate <= upper_bound)

```

```

aco_data <- aco_data %>%
  mutate(
    total_expenditures = rowSums(across(all_of(expenditure_columns)), na.rm = TRUE)
  ) %>%
  mutate(across(
    all_of(expenditure_columns),
    ~ round(.x / total_expenditures * 100, 2),
    .names = "{.col}_pct"
  ))

### Convert savings rate to percentage
aco_data <- aco_data %>%
  mutate(sav_rate = sav_rate * 100)

### Prepare panel and model datasets
fe_data <- pdata.frame(aco_data, index = c("aco_num", "year"))
glm_data <- as.data.frame(aco_data)

### Create binary savings indicator
glm_data$saved <- ifelse(glm_data$sav_rate > 0, 1, 0)

### Ensure 'year' is treated as a factor
glm_data$year <- as.factor(glm_data$year)

```

Unadjusted Year Fixed-Effects on Expenditures

Unadjusted FE

```
## Unadjusted Year Fixed-Effects on Expenditures

# Loop through each expenditure category (excluding outpatient)
unadjusted_fe <- lapply(
  expenditure_columns[expenditure_columns != "capann_opd"],
  function(var) {
    formula <- as.formula(paste(var, "~ factor(year) + factor(aco_num)")
    lm(formula, data = fe_data)
  }
)

# Rename models for table clarity
names(unadjusted_fe) <- c(
  "Inpatient Expenditures ($)",
  "Skilled Nursing Facilities ($)",
  "Physician Services ($)",
  "Ambulance Services ($)",
  "Home Health ($)",
  "Durable Medical Equipment ($)",
  "Hospice ($)"
)

# Clean coefficient labels for readability
coef_labels <- c(
  "factor(year)2015" = "Year: 2015",
  "factor(year)2016" = "Year: 2016",
  "factor(year)2017" = "Year: 2017"
)

# Output results as a clean HTML table
modelsummary(
  unadjusted_fe,
  output = "html",
  title = "Unadjusted Year Fixed-Effects on Expenditures (2014–2017)",
  stars = TRUE,
  coef_map = coef_labels,
  gof_omit = "IC|Log|Adj",
  statistic = "{std.error}",
  estimate = "{estimate}{stars}"
)
```

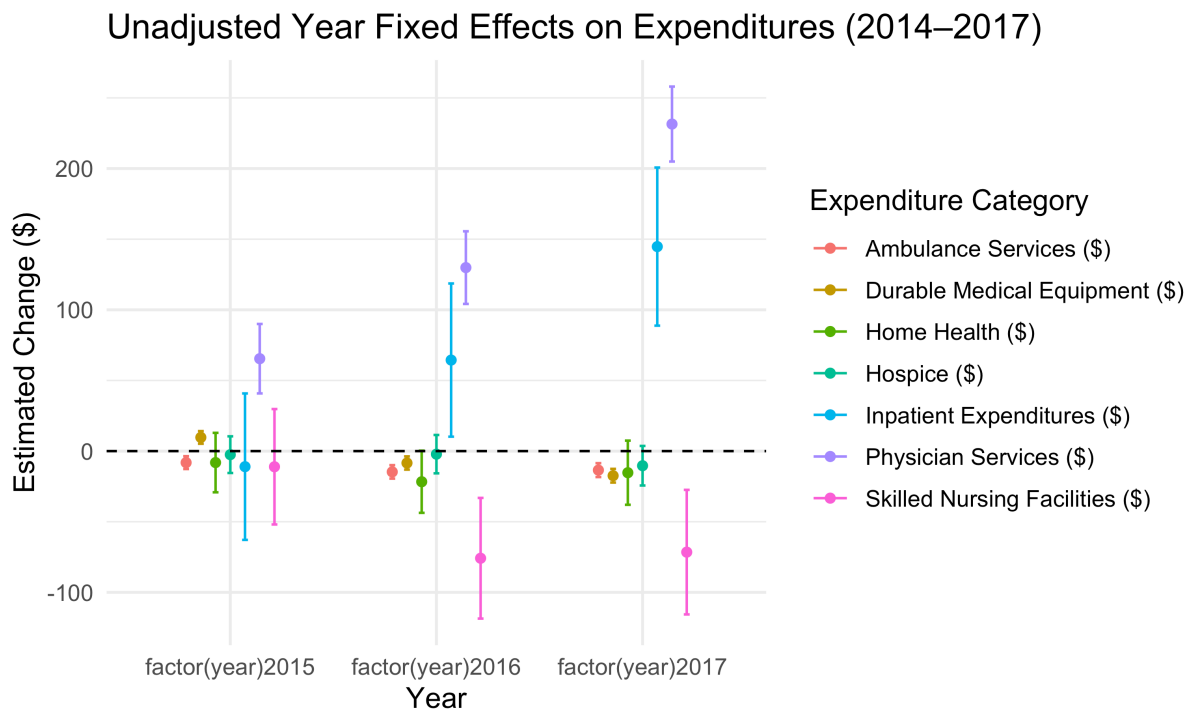
Unadjusted Year Fixed-Effects on Expenditures (2014–2017)

	Inpatient Expenditures (\$)	Skilled Nursing Facilities (\$)	Physician Services (\$)	Ambulance Services (\$)	Home Health (\$)	Durable Medica Equipment (\$)
Year: 2015	–11.016	–11.078	65.418***	–8.145***	–8.116	9.678***
	(26.449)	(20.836)	(12.541)	(2.321)	(10.736)	(2.286)
Year: 2016	64.429*	–75.837***	129.880***	–14.707***	–21.691+	–8.421***
	(27.661)	(21.791)	(13.115)	(2.427)	(11.228)	(2.391)
Year: 2017	144.750***	–71.519**	231.491***	–13.462***	–15.289	–17.417***
	(28.552)	(22.493)	(13.538)	(2.505)	(11.590)	(2.468)
Num.Obs.	1570	1569	1570	1570	1570	1570
R2	0.939	0.915	0.981	0.919	0.938	0.868
F	23.570	16.616	78.818	17.395	23.302	10.181
RMSE	254.78	200.77	120.80	22.36	103.42	22.02

```
# Coefficient Plot of year effects - all FE models
library(broom)
library(dplyr)
library(ggplot2)

# Combine and tidy results
tidy_unadj <- lapply(names(unadjusted_fe), function(name) {
  broom::tidy(unadjusted_fe[[name]]) %>%
    filter(grepl("factor\\(year\\)", term)) %>%
    mutate(model = name)
}) %>%
  bind_rows()

# Plot
ggplot(tidy_unadj, aes(x = term, y = estimate, color = model)) +
  geom_point(position = position_dodge(width = 0.5)) +
  geom_errorbar(aes(ymin = estimate - 1.96 * std.error,
                    ymax = estimate + 1.96 * std.error,
                    width = 0.2, position = position_dodge(width = 0.5))) +
  geom_hline(yintercept = 0, linetype = "dashed") +
  labs(
    title = "Unadjusted Year Fixed Effects on Expenditures (2014–2017)",
    x = "Year",
    y = "Estimated Change ($)",
    color = "Expenditure Category"
  ) +
  theme_minimal(base_size = 12)
```



Fixed-Effects Regression on Savings Rate

Savings Rate Model 1

```

### Savings Rate Model 1 – Full Fixed Effects Regression on Savings Rate (polished)

savings_rate_model1 <- plm(
  sav_rate ~ capann_inp_all_pct + capann_hsp_pct + capann_snf_pct +
    capann_pb_pct + capann_ambpay_pct + capann_hha_pct + capann_dme_pct,
  data = fe_data,
  model = "within"
)

# Clustered standard errors at the ACO level (match Model 2)
library(lmtest); library(sandwich)
se_model1 <- vcovHC(savings_rate_model1, type = "HCl", cluster = "group")

FE_map1 <- c(
  capann_inp_all_pct = "Inpatient Expenditures (%)",
  capann_hsp_pct = "Hospice Expenditures (%)",
  capann_snf_pct = "SNF Expenditures (%)",
  capann_pb_pct = "Physician Expenditures (%)",
  capann_ambpay_pct = "Ambulance Expenditures (%)",
  capann_hha_pct = "Home Health Expenditures (%)",
  capann_dme_pct = "DME Expenditures (%)"
)

modelsummary(
  savings_rate_model1,
  vcov = se_model1,
  coef_map = FE_map1,
  stars = TRUE,
  title = "Fixed-Effects Regression on Savings Rate (Full Model)",
  gof_omit = "IC|Log|Adj|Within",
  estimate = "{estimate}{stars}",
  statistic = "({std.error})",
  output = "html"
)

```

Fixed-Effects Regression on Savings Rate
(Full Model)

	(1)
Inpatient Expenditures (%)	−0.550***
	(0.159)
Hospice Expenditures (%)	−0.350
	(0.240)
SNF Expenditures (%)	−0.357**
	(0.128)
Physician Expenditures (%)	0.012
	(0.135)
Ambulance Expenditures (%)	−0.760
	(0.514)
Home Health Expenditures (%)	−0.423*
	(0.166)
DME Expenditures (%)	−0.332
	(0.338)
Num.Obs.	1569
R2	0.085

RMSE	2.39
Std.Errors	Custom

Savings Rate Model 2

```
### Savings Rate Model 2 – Reduced Fixed Effects Regression on Savings Rate

savings_rate_model2 <- plm(
  sav_rate ~ capann_inp_all_pct + capann_hsp_pct + capann_snf_pct +
    capann_pb_pct + capann_hha_pct,
  data = fe_data,
  model = "within"
)

# Clustered standard errors by ACO
library(lmtest)
library(sandwich)
se_model2 <- vcovHC(savings_rate_model2, type = "HCl", cluster = "group")

# Summarize with clustered SEs
coeftest(savings_rate_model2, vcov = se_model2)

##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## capann_inp_all_pct -0.498527   0.149706 -3.3301 0.0009017 ***
## capann_hsp_pct     -0.310974   0.231520 -1.3432 0.1795322
## capann_snf_pct     -0.369178   0.124118 -2.9744 0.0030097 **
## capann_pb_pct       0.040402   0.131701  0.3068 0.7590843
## capann_hha_pct     -0.379830   0.163907 -2.3173 0.0206966 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Coefficient name map for clarity in tables
FE_map2 <- c(
  "capann_inp_all_pct" = "Inpatient Expenditures (%)",
  "capann_hsp_pct"     = "Hospice Expenditures (%)",
  "capann_snf_pct"     = "SNF Expenditures (%)",
  "capann_pb_pct"      = "Physician Expenditures (%)",
  "capann_hha_pct"     = "Home Health Expenditures (%)"
)

# Polished output table for your HTML report
modelsummary(
  savings_rate_model2,
  vcov      = se_model2,
  coef_map = FE_map2,
  stars     = TRUE,
  output    = "html",
  title     = "Fixed-Effects Regression on Savings Rate (Reduced Model)",
  gof_omit  = "IC|Log|Adj|Within",
  estimate  = "{estimate}{stars}",
  statistic = "({std.error})"
)
```

Fixed-Effects Regression on Savings Rate (Reduced Model)	
(1)	
Inpatient Expenditures (%)	−0.499***
(0.150)	

Hospice Expenditures (%)	−0.311
	(0.232)
SNF Expenditures (%)	−0.369**
	(0.124)
Physician Expenditures (%)	0.040
	(0.132)
Home Health Expenditures (%)	−0.380*
	(0.164)
Num.Obs.	1569
R ²	0.082
RMSE	2.40
Std.Errors	Custom

Logit model

Logit Model 1

```

## =====
## Logit models: saved ~ composition + FE
## =====

# Rebuild a fresh dataset so chunk order cannot bite us
glm_data <- as.data.frame(aco_data)

# Outcome and fixed effects
glm_data$saved <- as.integer(glm_data$sav_rate > 0) # 1 if saved
glm_data$year <- factor(glm_data$year)
glm_data$aco_num <- factor(glm_data$aco_num)

# Sanity checks
needed_vars <- c(
  "saved",
  "capann_inp_all_pct", "capann_hsp_pct", "capann_snf_pct",
  "capann_pb_pct", "capann_ambpay_pct", "capann_hha_pct", "capann_dme_pct",
  "aco_num", "year"
)
stopifnot(all(needed_vars %in% names(glm_data)))

# Packages
library(fixest)
library(modelsummary)
library(broom)
library(dplyr)

# Helper for nice labels
coef_map_logit <- c(
  capann_inp_all_pct = "Inpatient Expenditures (%)",
  capann_hsp_pct = "Hospice Expenditures (%)",
  capann_snf_pct = "SNF Expenditures (%)",
  capann_pb_pct = "Physician Expenditures (%)",
  capann_ambpay_pct = "Ambulance Expenditures (%)",
  capann_hha_pct = "Home Health Expenditures (%)",
  capann_dme_pct = "DME Expenditures (%)"
)

# Fit logits with ACO and year fixed effects
logit_full <- feglm(
  saved ~ capann_inp_all_pct + capann_hsp_pct + capann_snf_pct +
    capann_pb_pct + capann_ambpay_pct + capann_hha_pct + capann_dme_pct |
    aco_num + year,
  data = glm_data,
  family = binomial()
)

logit_reduced <- feglm(
  saved ~ capann_inp_all_pct + capann_hsp_pct + capann_snf_pct +
    capann_pb_pct + capann_hha_pct |
    aco_num + year,
  data = glm_data,
  family = binomial()
)

# Polished table with clustered SEs by ACO
modelsummary(
  list("Logit: Full" = logit_full, "Logit: Reduced" = logit_reduced),
  coef_map = coef_map_logit,
  vcov = ~ aco_num, # cluster at ACO
  gof_omit = "IC|Log|Adj|Within",
  estimate = "{estimate}{stars}",
  statistic = "({std.error})",
  stars = TRUE,
  output = "html",
  title = "Probability of Achieving Savings: Fixed-Effects Logits"
)

```


	Logit: Full	Logit: Reduced
Inpatient Expenditures (%)	-0.559*	-0.541*
	(0.235)	(0.227)
Hospice Expenditures (%)	-0.302	-0.304
	(0.392)	(0.383)
SNF Expenditures (%)	-0.183	-0.222
	(0.173)	(0.168)
Physician Expenditures (%)	0.228	0.238
	(0.206)	(0.207)
Ambulance Expenditures (%)	-0.689	
	(0.842)	
Home Health Expenditures (%)	-0.373	-0.344
	(0.229)	(0.210)
DME Expenditures (%)	0.406	
	(0.647)	
Num.Obs.	762	762
R2	0.192	0.190
RMSE	0.43	0.43
Std.Errors	by: aco_num	by: aco_num
FE: aco_num	X	X
FE: year	X	X

```
# Odds ratios for the reduced model (cleaner to read)
or_tbl <- tidy(logit_reduced, conf.int = TRUE, cluster = "aco_num") |>
  filter(term %in% names(coef_map_logit)) |>
  mutate(
    term = recode(term, !!!coef_map_logit),
    OR   = exp(estimate),
    LCL  = exp(conf.low),
    UCL  = exp(conf.high)
  ) |>
  select(Predictor = term, `Odds ratio` = OR, `95% LCL` = LCL, `95% UCL` = UCL)

or_tbl
```

Predictor <chr>	Odds ratio <dbl>	95% LCL <dbl>	95% UCL <dbl>
Inpatient Expenditures (%)	0.5820866	0.3730611	0.9082288
Hospice Expenditures (%)	0.7380627	0.3486993	1.5621954
SNF Expenditures (%)	0.8006910	0.5760277	1.1129779
Physician Expenditures (%)	1.2692586	0.8467469	1.9025962
Home Health Expenditures (%)	0.7089243	0.4697636	1.0698438
5 rows			

Logit Model 2

```

### Logit Model 2 – Reduced spec with FE and clustered SEs

# Rebuild clean data for safety
glm_data <- as.data.frame(aco_data)
glm_data$saved <- as.integer(glm_data$sav_rate > 0)
glm_data$year <- factor(glm_data$year)
glm_data$aco_num <- factor(glm_data$aco_num)

# Packages
library(fixest)
library(modelsummary)
library(broom)
library(dplyr)

# Reduced specification to ease collinearity
logit_model2 <- feglm(
  saved ~ capann_inp_all_pct + capann_hsp_pct + capann_snf_pct +
    capann_pb_pct + capann_hha_pct |
    aco_num + year,
  data = glm_data,
  family = binomial()
)

# Pretty labels
logit_map2 <- c(
  capann_inp_all_pct = "Inpatient Expenditures (%)",
  capann_hsp_pct = "Hospice Expenditures (%)",
  capann_snf_pct = "SNF Expenditures (%)",
  capann_pb_pct = "Physician Expenditures (%)",
  capann_hha_pct = "Home Health Expenditures (%)"
)

# Portfolio-ready table with ACO-clustered SEs
modelsummary(
  list("Logit: Reduced" = logit_model2),
  coef_map = logit_map2,
  vcov = ~ aco_num, # cluster at ACO
  estimate = "{estimate}{stars}",
  statistic = "({std.error})",
  stars = TRUE,
  gof_omit = "IC|Log|Adj|Within",
  output = "html",
  title = "Probability of Achieving Savings: Reduced Fixed-Effects Logit"
)

```

Probability of Achieving Savings: Reduced Fixed-Effects Logit

	Logit: Reduced
Inpatient Expenditures (%)	−0.541*
	(0.227)
Hospice Expenditures (%)	−0.304
	(0.383)
SNF Expenditures (%)	−0.222
	(0.168)
Physician Expenditures (%)	0.238
	(0.207)
Home Health Expenditures (%)	−0.344
	(0.210)

Num.Obs.	762
R2	0.190
RMSE	0.43
Std.Errors	by: aco_num
FE: aco_num	X
FE: year	X

```
# Odds ratios for quick interpretation
or_tbl <- tidy(logit_model2, conf.int = TRUE, cluster = "aco_num") |>
  filter(term %in% names(logit_map2)) |>
  mutate(
    Predictor = recode(term, !!!logit_map2),
    `Odds ratio` = exp(estimate),
    `95% LCL`    = exp(conf.low),
    `95% UCL`    = exp(conf.high)
  ) |>
  select(Predictor, `Odds ratio`, `95% LCL`, `95% UCL`)

or_tbl
```

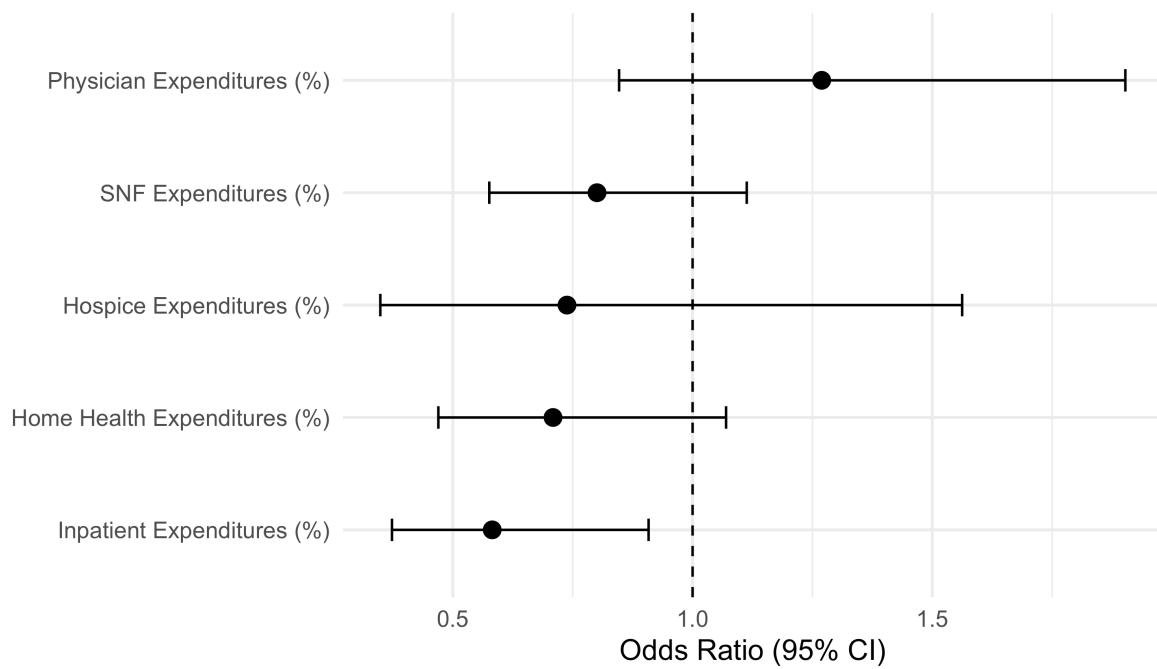
Predictor <chr>	Odds ratio <dbl>	95% LCL <dbl>	95% UCL <dbl>
Inpatient Expenditures (%)	0.5820866	0.3730611	0.9082288
Hospice Expenditures (%)	0.7380627	0.3486993	1.5621954
SNF Expenditures (%)	0.8006910	0.5760277	1.1129779
Physician Expenditures (%)	1.2692586	0.8467469	1.9025962
Home Health Expenditures (%)	0.7089243	0.4697636	1.0698438
5 rows			

```
library(ggplot2)
library(broom)

# Plot odds ratios for the reduced model
or_data <- tidy(logit_model2, conf.int = TRUE) |>
  filter(term %in% names(logit_map2)) |>
  mutate(
    OR = exp(estimate),
    LCL = exp(conf.low),
    UCL = exp(conf.high),
    term = recode(term, !!!logit_map2)
  )

ggplot(or_data, aes(x = reorder(term, OR), y = OR)) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = LCL, ymax = UCL), width = 0.2) +
  geom_hline(yintercept = 1, linetype = "dashed") +
  coord_flip() +
  labs(
    title = "Odds of Achieving Savings by Spending Composition (Reduced Logit Model)",
    x = "",
    y = "Odds Ratio (95% CI)"
  ) +
  theme_minimal(base_size = 12)
```

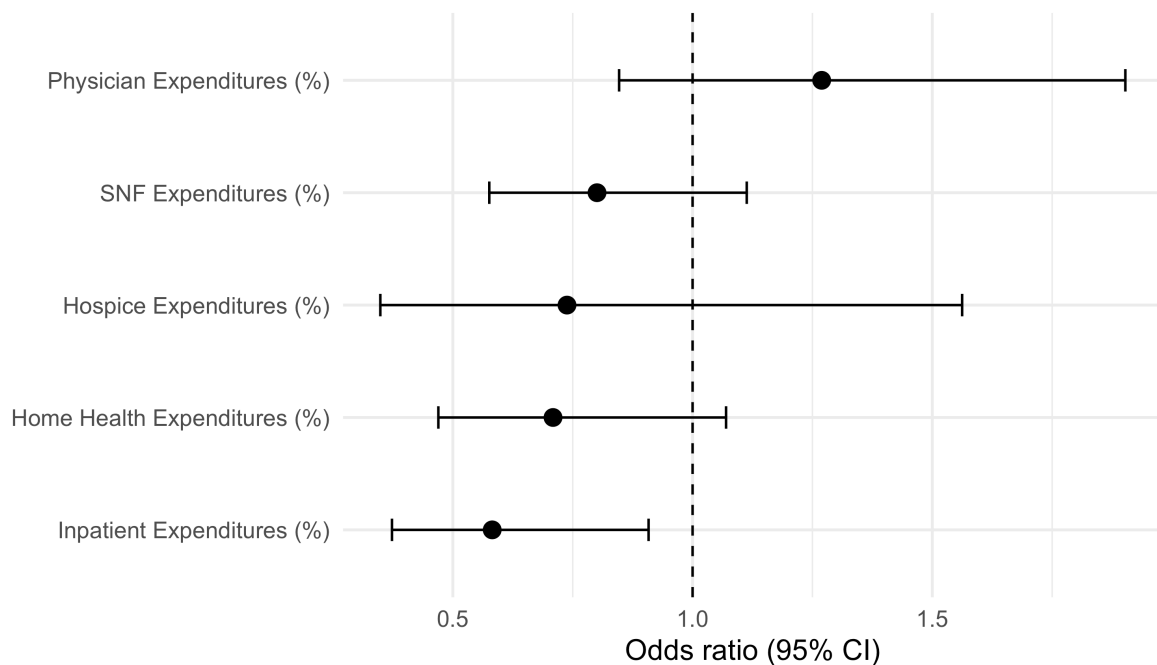
Odds of Achieving Savings by Spending Composition (



```
library(ggplot2)

ggplot(or_tbl, aes(x = reorder(Predictor, `Odds ratio`), y = `Odds ratio`)) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = `95% LCL`, ymax = `95% UCL`), width = 0.2) +
  geom_hline(yintercept = 1, linetype = "dashed") +
  coord_flip() +
  labs(
    title = "Odds of Achieving Savings by Spending Composition",
    x = "",
    y = "Odds ratio (95% CI)"
  ) +
  theme_minimal(base_size = 12)
```

Odds of Achieving Savings by Spending Composition



Combined Results Table

```

## Combined Results Table – FE + Logit

# Nice labels for coefficients (match your cleaned names)
capann_vars <- c(
  capann_inp_all_pct = "Percent Inpatient",
  capann_hsp_pct     = "Percent Hospice",
  capann_snf_pct     = "Percent SNF",
  capann_pb_pct      = "Percent Physician",
  capann_ambpay_pct  = "Percent AmbPay",
  capann_hha_pct     = "Percent HHA",
  capann_dme_pct     = "Percent DME"
)

# Make sure you have clustered vcovs for the FE models
# If you already computed these earlier, reuse them; otherwise:
library(lmtest); library(sandwich)
se_fe_full    <- vcovHC(savings_rate_model1, type = "HC1", cluster = "group")
se_fe_reduced <- vcovHC(savings_rate_model2, type = "HC1", cluster = "group")

# For fixest logits we can pass a formula to cluster by ACO
vcov_list <- list(se_fe_full, se_fe_reduced, ~ aco_num, ~ aco_num)

# Assemble models in the order you want to show
model_list <- list(
  "FE: Savings Rate 1 (Full)"   = savings_rate_model1,
  "FE: Savings Rate 2 (Reduced)" = savings_rate_model2,
  "Logit: Full"                 = logit_full,
  "Logit: Reduced"              = logit_model2
)

library(modelsummary)

# HTML for your portfolio page
msummary(
  model_list,
  vcov      = vcov_list,
  coef_map  = capann_vars,
  estimate  = "{estimate}{stars}",
  statistic = "{std.error}",
  stars     = TRUE,
  gof_omit  = "IC|Log|Adj|Within|RMSE",
  title     = "Spending Composition and ACO Savings: Fixed Effects and Logit Models",
  notes     = "FE models report coefficients on savings rate with ACO-clustered SEs. Logit models
are fixed-effects logits with ACO and year FEs; SEs clustered by ACO. Coefficients for logits are
log-odds.",
  output    = "outputs/combined_results.html"
)

# Optional DOCX export for applications
msummary(
  model_list,
  vcov      = vcov_list,
  coef_map  = capann_vars,
  estimate  = "{estimate}{stars}",
  statistic = "{std.error}",
  stars     = TRUE,
  gof_omit  = "IC|Log|Adj|Within|RMSE",
  title     = "Spending Composition and ACO Savings: Fixed Effects and Logit Models",
  notes     = "FE models report coefficients on savings rate with ACO-clustered SEs. Logit models
are fixed-effects logits with ACO and year FEs; SEs clustered by ACO. Coefficients for logits are
log-odds.",
  output    = "outputs/combined_results.docx"
)

```