Epidemiology : Poster Presentation Assignment

Boaz Baliejukia

2024-10-19

Background

Low back pain (LBP) is a widespread condition that affects individuals across all age groups. It is inherently subjective, as it cannot be observed directly and is only identifiable through self-report by those experiencing it(1). Globally, it is estimated that approximately 619 million people live with LBP(2). Diagnosing the underlying cause remains a complex and often debated process, as multiple factors—both physical and behavioral—may contribute to its onset(3). Among known lifestyle-related risk factors, smoking and elevated body mass index (BMI) have been consistently associated with increased likelihood of developing LBP and seeking healthcare for it(4). Gaining a clearer understanding of the relationship between excess body weight and LBP is critical for informing prevention strategies.

Research Question: **Are adults with a body mass index (BMI) $\geq$ 30 kg/m² at greater risk of developing low back pain compared to those with lower BMI?**

To explore this question, we have been asked to simulate a prospective cohort study as part of an epidemiology assessment. The fictional dataset includes 2,038 adults (aged 25–75) from NHS Lothian, selected via a two-stage cluster sampling method from 10 out of 20 general practices. After baseline screening, individuals who already had low back pain (n = 935) were excluded. Participants' BMI was calculated using standardized measurements of height and weight, and they were categorized into two groups: BMI $\geq$ 30 kg/m² (n = 66) and BMI < 30 kg/m² (n = 1,972). The cohort was followed over one year to identify new cases of LBP.

Note: This project was completed for educational purposes and uses simulated data.

Loading packages

```r
library(tidyverse)
library(knitr)
```

Let's upload the dataset

```r
sample <- read.csv("Poster Presentation.csv")
head(sample, 15) %>%
  kable(align = "c")
```

id	sex	age	bmi_cat	bp_baseline	lost	bp_followup
6	0	52	2	0	0	0
9	0	39	2	0	0	0
13	0	30	3	1	NA	NA
15	0	50	1	0	0	0
18	0	60	1	0	0	0
19	0	37	1	0	0	0
23	1	51	2	0	0	0

id	sex	age	bmi_cat	bp_baseline	lost	bp_followup
24	0	39	1	0	0	1
30	0	50	2	0	0	0
32	1	58	2	0	1	NA
40	0	37	2	0	0	0
41	1	53	2	0	0	1
42	0	50	3	0	0	1
43	1	60	2	0	0	0
44	0	34	3	0	0	1

```
glimpse(sample)
```

```
## Rows: 2,973
## Columns: 7
## $ id          <int> 6, 9, 13, 15, 18, 19, 23, 24, 30, 32, 40, 41, 42, 43, 44, ~
## $ sex         <int> 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0~
## $ age         <int> 52, 39, 30, 50, 60, 37, 51, 39, 50, 58, 37, 53, 50, 60, 34~
## $ bmi_cat     <int> 2, 2, 3, 1, 1, 1, 2, 1, 2, 2, 2, 2, 3, 2, 3, 3, 1, 2, 1, 2~
## $ bp_baseline <int> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0~
## $ lost        <int> 0, 0, NA, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, NA, 0,~
## $ bp_followup <int> 0, 0, NA, 0, 0, 0, 0, 1, 0, NA, 0, 1, 1, 0, 1, 0, 0, NA, 0~
```

We are now cleaning data

Let's recode variables

```
sample$id <- as.character(sample$id)
sample$sex <- factor(sample$sex, levels = c(0, 1), labels = c("Male", "Female"))
sample$bmi_cat <- factor(sample$bmi_cat, levels = c(1, 2, 3, 4, 5),
                         labels = c("Underweight", "Normal", "Overweight",
                                    "Obesity", "Severe obesity"))
sample$bp_baseline <- factor(sample$bp_baseline, levels = c(0, 1),
                             labels = c("No", "Yes"))
sample$lost <- factor(sample$lost, levels = c(0, 1),
                             labels = c("No", "Yes"))
sample$bp_followup <- factor(sample$bp_followup, levels = c(0, 1),
                             labels = c("No", "Yes"))

head(sample, 10) %>%
  kable(align = "c")
```

id	sex	age	bmi_cat	bp_baseline	lost	bp_followup
6	Male	52	Normal	No	No	No
9	Male	39	Normal	No	No	No
13	Male	30	Overweight	Yes	NA	NA
15	Male	50	Underweight	No	No	No
18	Male	60	Underweight	No	No	No
19	Male	37	Underweight	No	No	No
23	Female	51	Normal	No	No	No
24	Male	39	Underweight	No	No	Yes
30	Male	50	Normal	No	No	No
32	Female	58	Normal	No	Yes	NA

Let's get our population at risk and call it "cohort": adults without LBP at the beginning of the study (bp_baseline = No)

```
cohort <- sample[sample$bp_baseline == "No", ]
head(cohort, 10) %>%
  kable(align = "c")
```

	id	sex	age	bmi_cat	bp_baseline	lost	bp_followup
1	6	Male	52	Normal	No	No	No
2	9	Male	39	Normal	No	No	No
4	15	Male	50	Underweight	No	No	No
5	18	Male	60	Underweight	No	No	No
6	19	Male	37	Underweight	No	No	No
7	23	Female	51	Normal	No	No	No
8	24	Male	39	Underweight	No	No	Yes
9	30	Male	50	Normal	No	No	No
10	32	Female	58	Normal	No	Yes	NA
11	40	Male	37	Normal	No	No	No

```
cat <- cohort %>%
    group_by(bmi_cat) %>%
    summarise(
        count = n(),
        mean_age = mean(age, na.rm = TRUE),
        sd_age = sd(age, na.rm = TRUE),
        male_percentage = sum(sex == "Male") / n() * 100,
        female_percentage = sum(sex == "Female") / n() * 100
        )
cat %>%
  kable(align = "c")
```

Table 1: Baseline characteristics according to BMI categories

bmi_cat	count	mean_age	sd_age	male_percentage	female_percentage
Underweight	745	47.03490	12.56077	65.90604	34.09396
Normal	695	47.31223	12.68547	60.14388	39.85612
Overweight	532	47.75188	12.19145	63.53383	36.46617
Obesity	66	49.56061	12.92122	62.12121	37.87879

```
graph1 <- cohort %>%
    drop_na() %>%
    ggplot(aes(bmi_cat, fill = sex)) +
    geom_bar(position = "dodge", alpha = 0.8) +
    theme_classic(base_size = 13) +
  labs(title = "Participants by BMI groups and gender at baseline",
       x = "BMI categories", y = "Number of participants")

graph1
```

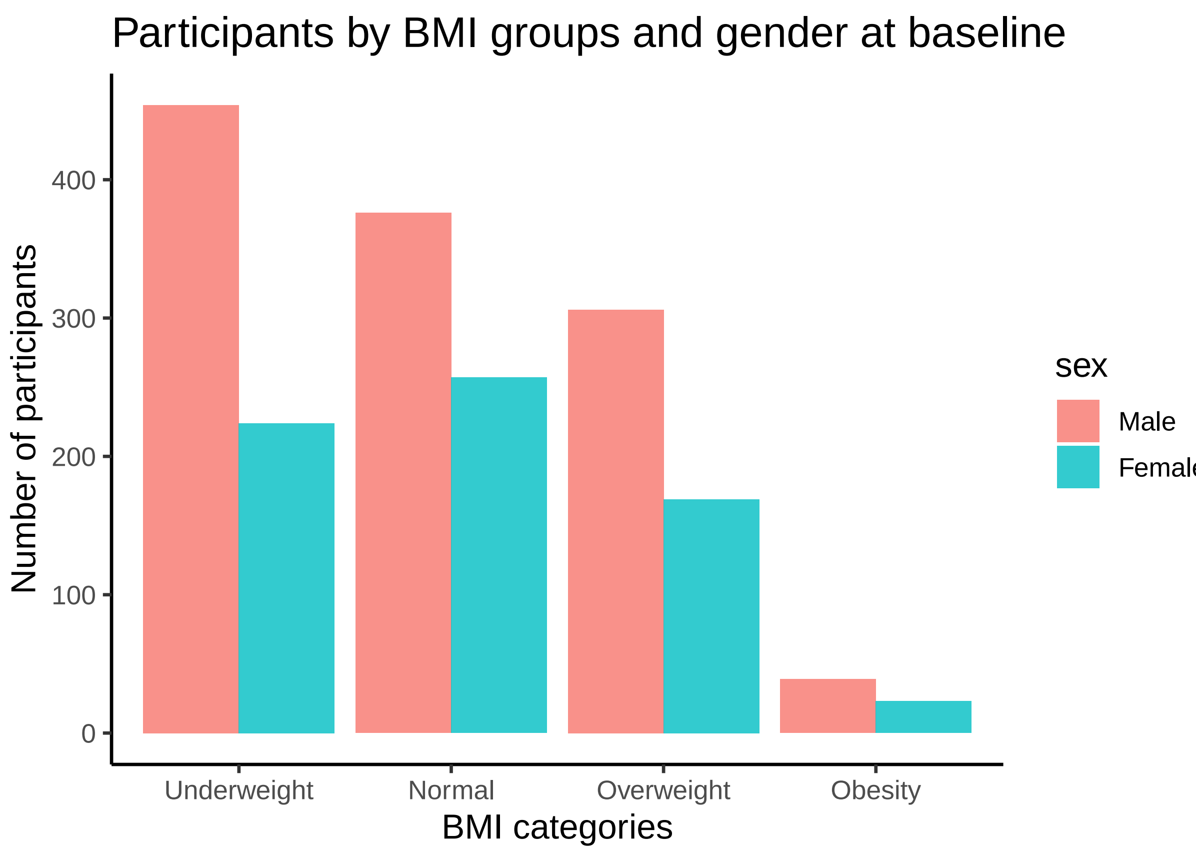

Figure 1: Barplot

We're now splitting our cohort into 2 groups: exposed (obesity) and non-exposed

```
exposed <- cohort[cohort$bmi_cat =="Obesity", ]
non_exposed <- cohort[!cohort$bmi_cat =="Obesity", ]

nrow(exposed)
```

```
## [1] 66
```

```
nrow(non_exposed)
```

```
## [1] 1972
```

```
demo_expo <- exposed %>%
    summarise(
        count = n(),
        mean_age = mean(age, na.rm = TRUE),
        sd_age = sd(age, na.rm = TRUE),
        male_percentage = sum(sex == "Male") / n() * 100,
        female_percentage = sum(sex == "Female") / n() * 100
    )
demo_expo %>%
  kable(align = "c")
```

Baseline Demographic characteristics of participants:

count	mean_age	sd_age	male_percentage	female_percentage
66	49.56061	12.92122	62.12121	37.87879

```
demo_non_expo <- non_exposed %>%
    summarise(
        count = n(),
        mean_age = mean(age, na.rm = TRUE),
        sd_age = sd(age, na.rm = TRUE),
        male_percentage = sum(sex == "Male") / n() * 100,
        female_percentage = sum(sex == "Female") / n() * 100
    )
demo_non_expo %>%
  kable(align = "c")
```

count	mean_age	sd_age	male_percentage	female_percentage
1972	47.32606	12.50358	63.23529	36.76471

Result

Attrition bias (lost to follow up):

```
lost_expo <- sum(is.na(exposed$bp_followup))
lost_non_expo <- sum(is.na(non_exposed$bp_followup))

lost_expo
```

```
## [1] 4
```

```
lost_non_expo
```

```
## [1] 186
```

Let's create a 2x2 table

Considering : a = exposed with outcome b = exposed without outcome c = non-exposed with outcome d = non-exposed without outcome

```
a <- length(which(exposed$bp_followup == "Yes"))
b <- length(which(exposed$bp_followup == "No"))
c <- length(which(non_exposed$bp_followup == "Yes"))
d <- length(which(non_exposed$bp_followup == "No"))

a
```

```
## [1] 25
```

```
b
```

```
## [1] 37
```

```
c
```

```
## [1] 625
```

```
d
```

```
## [1] 1161
```

We are now creating a 2x2 matrix :

```r
table <- matrix(c(a, b, c, d), nrow = 2, byrow = TRUE)
table <- cbind(table, c(a + b, c + d))
colnames(table) <- c("Low back pain", "Without LBP", "Total")
rownames(table) <- c("Obesity", "Without Obesity")

table %>%
  kable(align = "c")
```

	Low back pain	Without LBP	Total
Obesity	25	37	62
Without Obesity	625	1161	1786

Let's compute the Risk Ratio

```r
lbp_expo <- a/(a+b)

lbp_expo
```

Risk of outcome in exposed and non-exposed groups

```
## [1] 0.4032258
```

```r
lbp_non_exp <- c/(c+d)

lbp_non_exp
```

```
## [1] 0.349944
```

```r
# Risk Ratio and the Population Attributable Risk
RR <- lbp_expo/lbp_non_exp

RR
```

```
## [1] 1.152258
```

```r
PAR = 0.48
```

Individuals with BMI $\geq$ 30 had **1.15** times greater risk of LBP comparing to those with BMI <30 kg/m2. And **0.48%** of LBP cases is attributable to BMI $\geq$ 30.

Table 2: Summary table

```r
table1 <- data.frame(
  characteristics = c("Number of participants",
                      "Mean age",
                      "SD",
                      "Male",
                      "M %",
                      "Female",
                      "F %",
                      "Low Back Pain",
                      "LBP %",
                      "LBP M",
                      "LBPM %",
```

```
                      "LBP F",
                      "LBPF %"),
  Exposed = c(nrow(exposed),
                       round(mean(exposed$age), 2),
              round(sd(exposed$age), 2),
                       length(which(exposed$sex == "Male")),
              round(length(which(exposed$sex == "Male")) /
                      nrow(exposed) * 100, 2),
                       length(which(exposed$sex == "Female")),
              round(length(which(exposed$sex == "Female")) /
                      nrow(exposed) * 100, 2),
                       a, round(a/nrow(exposed) * 100, 2),
              length(which(exposed$sex == "Male" &
                            exposed$bp_followup == "Yes")),
              round(length(which(exposed$sex == "Male" &
                                  exposed$bp_followup == "Yes")) /
                      a * 100, 2),
              length(which(exposed$sex == "Female" &
                            exposed$bp_followup == "Yes")),
              round(length(which(exposed$sex == "Female" &
                                  exposed$bp_followup == "Yes")) /
                      a * 100, 2)),
  Non_exposed = c(nrow(non_exposed),
                       round(mean(non_exposed$age), 2),
                  round(sd(non_exposed$age), 2),
                       length(which(non_exposed$sex == "Male")),
                  round(length(which(non_exposed$sex == "Male")) /
                          nrow(non_exposed) * 100, 2),
                       length(which(non_exposed$sex == "Female")),
                  round(length(which(non_exposed$sex == "Female")) /
                          nrow(non_exposed) * 100, 2),
                       c, round(c/nrow(non_exposed) * 100, 2),
                  length(which(non_exposed$sex == "Male" &
                                non_exposed$bp_followup == "Yes")),
              round(length(which(non_exposed$sex == "Male" &
                                  non_exposed$bp_followup == "Yes")) /
                      c * 100, 2),
              length(which(non_exposed$sex == "Female" &
                            non_exposed$bp_followup == "Yes")),
              round(length(which(non_exposed$sex == "Female" &
                                  non_exposed$bp_followup == "Yes")) /
                      c * 100, 2)))

table1 %>%
  kable(align = "c")
```

characteristics	Exposed	Non_exposed
Number of participants	66.00	1972.00
Mean age	49.56	47.33
SD	12.92	12.50
Male	41.00	1247.00
M %	62.12	63.24
Female	25.00	725.00

characteristics	Exposed	Non_exposed
F %	37.88	36.76
Low Back Pain	25.00	625.00
LBP %	37.88	31.69
LBP M	13.00	378.00
LBPM %	52.00	60.48
LBP F	12.00	247.00
LBPF %	48.00	39.52

Table 3: age groups

```
table2 <- data.frame(Age_groups = c("25-35",
                          "36-45",
                          "46-55",
                          "56-65",
                          "66-75"),
           Number_of_participants = c(length(which(cohort$age < 36)),
                                      length(which(cohort$age > 35 &
                                                     cohort$age < 46)),
                                      length(which(cohort$age > 45 &
                                                     cohort$age < 56)),
                                      length(which(cohort$age > 55 &
                                                     cohort$age < 66)),
                                      length(which(cohort$age > 65))
                                      ),
           Number_of_cases_LBP = c(length(which(cohort$age < 36 &
                                                  cohort$bp_followup == "Yes")),
                                      length(which(cohort$age > 35 &
                                                     cohort$age < 46 &
                                                     cohort$bp_followup == "Yes")),
                                      length(which(cohort$age > 45 &
                                                     cohort$age < 56 &
                                                     cohort$bp_followup == "Yes")),
                                      length(which(cohort$age > 55 &
                                                     cohort$age < 66 &
                                                     cohort$bp_followup == "Yes")),
                                      length(which(cohort$age > 65 &
                                                     cohort$bp_followup == "Yes"))),
           Incidence_proportion = c(round(length(which(cohort$age < 36 &
                                                         cohort$bp_followup == "Yes")) /
                                      length(which(cohort$age < 36)) * 100, 2),
                                      round(length(which(cohort$age > 35 &
                                                           cohort$age < 46 &
                                                           cohort$bp_followup == "Yes")) /
                                      length(which(cohort$age > 35 &
                                                     cohort$age < 46)) * 100, 2),
                                      round(length(which(cohort$age > 45 &
                                                           cohort$age < 56 &
                                                           cohort$bp_followup == "Yes")) /
                                      length(which(cohort$age > 45 &
                                                     cohort$age < 56)) * 100, 2),
                                      round(length(which(cohort$age > 55 &
```

```
                                                                cohort$age < 66 &
                                                                cohort$bp_followup == "Yes")) /
                                              length(which(cohort$age > 55 &
                                                           cohort$age < 66)) * 100, 2),
                                              round(length(which(cohort$age > 65 &
                                                                cohort$bp_followup == "Yes")) /
                                              length(which(cohort$age > 65)) * 100, 2)))
table2 %>%
  kable(align = "c")
```

Age_groups	Number_of_participants	Number_of_cases_LBP	Incidence_proportion
25-35	434	47	10.83
36-45	489	150	30.67
46-55	506	162	32.02
56-65	461	143	31.02
66-75	148	148	100.00

Conclusion

This study demonstrates a clear association between high BMI and an elevated risk of developing LBP, particularly among men. These results highlight the importance of considering obesity as a modifiable risk factor in low back pain prevention strategies. Future studies should aim to clarify gender-specific risk profiles, integrate socioeconomic and lifestyle data and use extended follow-up for clearer long-term impact.

References

1. Manchikanti L. Epidemiology of low back pain. Pain Physician. 2000 Apr;3(2):167–92.
2. Ferreira ML, De Luca K, Haile LM, Steinmetz JD, Culbreth GT, Cross M. Global, regional, and national burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol. 2023;5(6):e316–29.
3. Knezevic NN, Candido KD, Vlaeyen JWS, Van Zundert J, Cohen SP. Low back pain. 2021;
4. Koyanagi A, Stickley A, Garin N, Miret M, Ayuso-Mateos JL, Leonardi M, et al. The association between obesity and back pain in nine countries: A cross-sectional study. BMC Public Health. 2015;15(1):1–9.