Independent Project 2

Bathsheba Aklilu

2025-04-06

Introduction

Breast cancer, like many other health issues, is characterized by inequitable survival rates, particularly among marginalized groups (Giaquinto et al., 2022). Disparities in health outcomes, such as age at diagnosis, quality of life, and overall survival, have been shown to vary significantly by race and ethnicity. These disparities may be exacerbated by the stressors and social determinants of health (SDOH) that individuals from marginalized communities face, in addition to their breast cancer diagnosis. Social determinants of health, such as income and access to healthcare, can contribute to increased inflammation and stress, which are often reflected in biomarkers like C-reactive protein (CRP) and cortisol levels (Antoni et al., 2020). Black and Latina/Hispanic women are disproportionately affected by these health disparities in breast cancer (Yedjou et al., 2019), yet they remain underrepresented in much of oncological research (Duma et al., 2018). C-reactive protein (CRP) is an important biomarker for inflammation, frequently utilized in clinical and epidemiological research to assess risk factors for cardiovascular disease, chronic inflammation, and overall health status. CRP levels can be influenced by various factors, including social determinants of health (SDOH) such as socioeconomic status, education, and access to healthcare. In particular, low socioeconomic status (SES) and limited access to healthcare are often associated with chronic stress, which can contribute to elevated CRP levels. While much of the research has focused on the role of SDOH in these disparities, few studies have explored how stress factors are related to these outcomes, particularly in a diverse cohort such as the NHANES dataset.

This analysis aims to investigate the relationship between CRP levels and social determinants of health among participants with history of breast cancer using the **NHANES 2021-2023 dataset**. We will focus on CRP as the dependent variable, exploring how social determinants of health such as income, education, and marital status impact inflammation. The National Health and Nutrition Examination Survey (NHANES) 2021-2023, includes detailed information on various health indicators, including CRP levels, income, and cancer history. We will examine CRP levels in relation to social determinants of health, including family income and poverty level, as well as demographic factors such as age and race/ethnicity. The primary statistical approaches utilized in this analysis include descriptive statistics, ANOVA, and regression modeling.

Research Question and Hypothesis

- **Research Question**: How do social determinants of health (SDOH) such as income, education, and marital status influence CRP levels among women with hisptory of breast cancer in the NHANES 2021-2023 cohort?
- **Hypothesis**: We hypothesize that individuals with lower income, lower education levels, and divorced marital status will have higher CRP levels, reflecting higher inflammation due to chronic stress and poor health.

Aim 1: Determine the association between social determinants of health (income, education, and marital status) and C-reactive protein by race/ethnicity among women with history of breast cancer.

Data Installation

```
# Install and load necessary packages
#install.packages("tidyverse")
library(tidyverse)
library(haven)

# loading data
crp_data <- read_xpt("HSCRP_L.xpt")
demo_data <- read_xpt("DEMO_L.xpt")
income_data <- read_xpt("INQ_L.xpt")
medical_data <- read_xpt("MCQ_L.xpt")
```

Data Cleaning and Preparation

```
# Clean and prepare CRP data (select necessary columns only)
crp_data_clean <- crp_data %>%
  select(SEQN, LBXHSCRP)

# Clean and prepare Demographics data (select necessary columns only)
demo_data_clean <- demo_data %>%
  select(SEQN, RIAGENDR, RIDAGEYR, RIDRETH1, INDFMPIR, DMDEDUC2, DMDMARTZ)

# Clean and prepare Income data (select necessary columns only)
income_data_clean <- income_data %>%
  select(SEQN, INDFMMPI)

# Clean and prepare Medical Conditions data (select necessary columns only)
breast_cancer_data <- medical_data %>%
  filter(MCQ230A == 14) %>%
  select(SEQN, MCQ230A)

# Merge the datasets by SEQN (participant id)
merged_data <- demo_data_clean %>%
  left_join(crp_data_clean, by = "SEQN") %>%
  left_join(income_data_clean, by = "SEQN") %>%
  inner_join(breast_cancer_data, by = "SEQN")

# Change coding of variables to characters
cleaned_data_labeled <- merged_data %>%
  mutate(
    RIDRETH1 = recode(RIDRETH1,
      `1` = "Mexican American",
      `2` = "Other Hispanic",
      `3` = "Non-Hispanic White",
      `4` = "Non-Hispanic Black",
      `5` = "Other Race - Including Multi-Racial"
    ),
    DMDMARTZ = recode(DMDMARTZ,
      `1` = "Married/Living with partner",
      `2` = "Widowed/Divorced/Separated",
```

```
      `3` = "Never married",
      `77` = "Refused",
      `99` = "Don't know"
    ),
    DMDEDUC2 = recode(DMDEDUC2,
      `1` = "Less than 9th grade",
      `2` = "9-11th grade (No diploma)",
      `3` = "High school graduate/GED",
      `4` = "Some college or AA degree",
      `5` = "College graduate or above",
      `7` = "Refused",
      `9` = "Don't know"
    ))%>%
  filter(
    DMDMARTZ != "Refused",
    DMDMARTZ != "Don't know",
    DMDEDUC2 != "Refused",
    DMDEDUC2 != "Don't know")
# View the cleaned and merged data
#merged_data

cleaned_compiled_all <- cleaned_data_labeled %>%
  drop_na()

cleaned_compiled_all
```

```
## # A tibble: 108 x 10
##      SEQN RIAGENDR RIDAGEYR RIDRETH1         INDFMPIR DMDEDUC2 DMDMARTZ LBXHSCRP
##     <dbl>    <dbl>    <dbl> <chr>               <dbl> <chr>    <chr>       <dbl>
##  1 130392        2       74 Non-Hispanic Wh~     3.04 College~ Married~     9.46
##  2 130407        2       73 Non-Hispanic Wh~     4.37 College~ Widowed~     0.24
##  3 130523        2       61 Non-Hispanic Wh~     5    College~ Married~     0.43
##  4 130826        2       79 Non-Hispanic Wh~     3.3  Some co~ Married~     8.74
##  5 131137        2       67 Other Hispanic       4.05 High sc~ Widowed~     0.71
##  6 131169        2       78 Non-Hispanic Wh~     5    Some co~ Widowed~     0.71
##  7 131342        2       61 Non-Hispanic Wh~     4.67 College~ Married~     0.27
##  8 131450        2       80 Non-Hispanic Wh~     3.68 College~ Widowed~     3.8
##  9 131509        2       45 Non-Hispanic Wh~     5    College~ Widowed~     0.4
## 10 131554        2       69 Non-Hispanic Wh~     5    College~ Widowed~    13.1
## # i 98 more rows
## # i 2 more variables: INDFMMPI <dbl>, MCQ230A <dbl>
```

Statistical Analysis

- **ANOVA/Kruskal-Wallis Test**: Depending on normality/ANOVA assumptions.

```
# ANOVA for Education Level
anova_edu <- aov(LBXHSCRP ~ DMDEDUC2, data = cleaned_compiled_all)
summary(anova_edu)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## DMDEDUC2      4  145.6   36.39   3.014 0.0214 *
```

```
## Residuals   103 1243.7   12.07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# ANOVA for Marital Status
anova_marital <- aov(LBXHSCRP ~ DMDMARTZ, data = cleaned_compiled_all)
summary(anova_marital)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## DMDMARTZ      2   29.5   14.77   1.141  0.324
## Residuals   105 1359.7   12.95
```

```
# ANOVA for Race/Ethnicity
anova_race <- aov(LBXHSCRP ~ RIDRETH1, data = cleaned_compiled_all)
summary(anova_race)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## RIDRETH1      4  101.2   25.30   2.024 0.0966 .
## Residuals   103 1288.1   12.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Kruskal-Wallis for Education
kruskal.test(LBXHSCRP ~ DMDEDUC2, data = cleaned_compiled_all)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  LBXHSCRP by DMDEDUC2
## Kruskal-Wallis chi-squared = 6.2027, df = 4, p-value = 0.1845
```

```
# Kruskal-Wallis for Marital status
kruskal.test(LBXHSCRP ~ DMDMARTZ, data = cleaned_compiled_all)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  LBXHSCRP by DMDMARTZ
## Kruskal-Wallis chi-squared = 0.062187, df = 2, p-value = 0.9694
```

```
# Kruskal-Wallis for Race/Ethnicity
kruskal.test(LBXHSCRP ~ RIDRETH1, data = cleaned_compiled_all)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  LBXHSCRP by RIDRETH1
## Kruskal-Wallis chi-squared = 3.1614, df = 4, p-value = 0.5312
```

- **Pearson's and Linear Regression**: Depending on normality.

```r
# Linear regression for CRP and Poverty Index
lm(LBXHSCRP ~ INDFMMPI, data = cleaned_compiled_all)
```

```
##
## Call:
## lm(formula = LBXHSCRP ~ INDFMMPI, data = cleaned_compiled_all)
##
## Coefficients:
## (Intercept)     INDFMMPI
##      5.6175      -0.6978
```

```r
# Pearsons for correlational coeffecient

# Correlation between CRP and family poverty index
cor.test(cleaned_compiled_all$LBXHSCRP, cleaned_compiled_all$INDFMMPI, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  cleaned_compiled_all$LBXHSCRP and cleaned_compiled_all$INDFMMPI
## t = -3.0666, df = 106, p-value = 0.002748
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.4501515 -0.1019887
## sample estimates:
##        cor
## -0.2854609
```

```r
# Correlation between CRP and income-to-poverty ratio
cor.test(cleaned_compiled_all$LBXHSCRP, cleaned_compiled_all$INDFMPIR, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  cleaned_compiled_all$LBXHSCRP and cleaned_compiled_all$INDFMPIR
## t = -3.6716, df = 106, p-value = 0.0003795
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.4935435 -0.1568819
## sample estimates:
##        cor
## -0.3358976
```

References: Aghaee, S., Allen, A., Ramirez, J., Shariff-Marco, S., Allen, L., DeRouen, M., . . . Gomez, S. L. (2022). Everyday discrimination and telomere length in a multiethnic cohort of breast cancer survivors. Ethnicity & Health, 27(3), 542–553. https://doi.org/10.1080/13557858.2020.1739231 Antoni, M. H., Lechner, S. C., Kilbourn, K. M., & Phillips, K. A. (2020). Behavioral, physical, and psychological predictors of cortisol and C-reactive protein in breast cancer survivors: A longitudinal study. Psycho-Oncology, 29(8), 1237-1245. https://doi.org/10.1002/pon.5397 Barnes, R. P., Fouquerel, E., & Opresko, P. L. (2019). The impact of oxidative DNA damage and stress on telomere homeostasis. Mechanisms of ageing and development, 177, 2 37–45. doi: 10.1016/j.mad.2018.03.013. PMID: 29604323; PMCID: PMC6162185.

Benetos, A., Okuda, K., Lajemi, M., Kimura, M., Thomas, F., Skurnick, J., Labat, C., Bean, K., & Aviv, A. (2001). Telomere length as an indicator of biological aging: The gender effect and relation with pulse pressure and pulse wave velocity. Hypertension, 37(2 Pt 2), 381–385. https://doi.org/10.1161/01.hyp.37.2.381 Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. Critical care (London, England), 7(6), 451–459. https://doi.org/10.1186/cc2401 Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys, 22(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x Cohen, S., Doyle, W. J., & Baum, A. (2006). Socioeconomic status is associated with stress hormones. Psychosomatic medicine, 68(3), 414–420. Doi: 10.1097/01.psy.0000221236.37158.b9. PMID: 16738073. Coughlin S. S. (2019). Social determinants of breast cancer risk, stage, and survival. Breast cancer research and treatment, 177(3), 537–548. https://doi.org/10.1007/s10549-019-05340-7 DeSantis, C. E., Ma, J., Sauer, A. G., Newman, L. A., & Jemal, A. (2017). Breast cancer statistics, 2017, racial disparity in mortality by state. CA: A Cancer Journal for Clinicians, 67(6), 439-448. https://doi.org/10.3322/caac.21412 Ding, Z., Mangino, M., Aviv, A., Spector, T., Durbin, R., & UK10K Consortium (2014). Estimating telomere length from whole genome sequence data. Nucleic acids research, 42(9), e75. https://doi.org/10.1093/nar/gku181 Duma, N., Vera Aguilera, J., Paludo, J., Haddox, C. L., Gonzalez Velez, M., Wang, Y., Leventakos, K., Hubbard, J. M., Mansfield, A. S., Go, R. S., & Adjei, A. A. (2018). Representation of Minorities and Women in Oncology Clinical Trials: Review of the Past 14 Years. Journal of oncology practice, 14(1), e1–e10. https://doi.org/10.1200/JOP.2017.025288 Farmery, J.H.R., Smith, M.L., NIHR BioResource - Rare Diseases. et al. Telomerecat: A ploidy-agnostic method for estimating telomere length from whole genome sequencing data. Sci Rep 8, 1300 (2018). https://doi.org/10.1038/s41598-017-14403-y Ferrer, A., Stephens, Z. D., & Kocher, J. A. (2023). Experimental and Computational Approaches to Measure Telomere Length: Recent Advances and Future Directions. Current hematologic malignancy reports, 18(6), 284–291. https://doi.org/10.1007/s11899-023-00717-4 Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, et al.. Breast cancer statistics, 2022. CA: Cancer J Clin (2022) 0:1–18. doi: 10.3322/caac.21754 Guo, L., Liu, S., Zhang, S., Chen, Q., Zhang, M., Quan, P., Lu, J., & Sun, X. (2015). C-reactive protein and risk of breast cancer: A systematic review and meta-analysis. Scientific reports, 5, 10508. https://doi.org/10.1038/srep10508 He, X.-Y., Gao, Y., Ng, D., Michalopoulou, E., George, S., Adrover, J. M., . . . & Egeblad, M. (2023). Chronic stress increases metastasis via neutrophil-mediated changes to the microenvironment. Nature, 616 (7956), 563–572. https://doi.org/10.1038/s41586-023-06020-3 Hopper, J.L., Dite, G.S., MacInnis, R.J. et al. Age-specific breast cancer risk by body mass index and familial risk: prospective family study cohort (ProF-SC). Breast Cancer Res 20, 132 (2018). https://doi.org/10.1186/s13058-018-1056-1 Islami, F., Ward, E. M., Sung, H., Cronin, K. A., Tangka, F. K. L., Sherman, R. L., Zhao, J., Anderson, R. N., Henley, S. J., Yabroff, K. R., Jemal, A., & Benard, V. B. (2021). Annual Report to the Nation on the Status of Cancer, Part 1: National Cancer Statistics. JNCI: Journal of the National Cancer Institute, 113(12), 1648–1669. https://doi.org/10.1093/jnci/djab131 Judasz, E., Lisiak, N., Kopczyński, P., Taube, M., & Rubiś, B. (2022). The Role of Telomerase in Breast Cancer's Response to Therapy. International journal of molecular sciences, 23(21), 12844. https://doi.org/10.3390/ijms232112844 Koliada, A. K., Krasnenkov, D. S., & Vaiserman, A. M. (2015, March 16). Telomeric aging: Mitotic clock or stress indicator? Frontiers. https://www.frontiersin.org/articles/10.3389/f gene.2015.00082/full Law, E., Girgis, A., Lambert, S., Sylvie, L., Levesque, J., & Pickett, H. (2016). Telomeres and Stress: Promising Avenues for Research in Psycho-Oncology. Asia-Pacific journal of oncology nursing, 3(2), 137–147. https://doi.org/10.4103/2347- 5625.182931 Li, Y., & Ma, L. (2022). Relationship between telomere length and the prognosis of breast cancer based on estrogen receptor status: A Mendelian randomization study. Frontiers in oncology, 12, 1024772. https://doi.org/10.3389/fonc.2022.1024772 Lin, J., & Epel, E. (2022). Stress and telomere shortening: Insights from cellular mechanisms. Ageing research reviews, 73, 101507. https://doi.org/10.1016/j.arr.2021.101507 Liu, S. H., Juster, R.-P., Dams-O'Connor, K., & Spicer, J. (2021). Allostatic load scoring using item response theory. Comprehensive Psychoneuroendocrinology, 5, 100025. https://doi.org/10.1016/j.cpnec.2020.100025 Lu, D., Palmer, J. R., Rosenberg, L., Shields, A. E., Orr, E. H., DeVivo, I., & Cozier, Y. C. (2019). Perceived racism in relation to telomere length among African American women in the Black Women's Health Study. Annals of epidemiology, 36, 33–39. Doi: 10.1016/j.annepidem.2019.06.003. PMID: 31387775; PMCID: PMC7048405. Mikkelsen, M. K., Lindblom, N. A. F., Dyhl-Polk, A., Juhl, C. B., Johansen, J. S., & Nielsen, D. (2022). Systematic review and meta-analysis of C-reactive protein as a biomarker in breast cancer. Critical 3

Reviews in Clinical Laboratory Sciences, 59(7), 480–500. https://doi.org/10.1080/10408363.2022.2050886 Nazmi, A., & Victora, C. G. (2007). Socioeconomic and racial/ethnic differentials of C-reactive protein levels: A systematic review of population-based studies - BMC Public Health. BioMed Central. https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-7-212#citeas Needham, B. L., Salerno, S., Roberts, E., Boss, J., Allgood, K. L., & Mukherjee, B. (2019). Do Black/White Differences in Telomere Length Depend on Socioeconomic Status? Biodemography and Social Biology, 65(4), 287. https://doi.org/10.1080/19485565.2020.1765734 Niu, B., Wu, J. X., Huang, X. L., Lei, S. F., & Deng, F. Y. (2024). Telomere Length Is a Driving Hallmark for Aging-Related Biochemical Hallmarks: Evidence From the Shared Genetic Effect and Causal Inference. The journals of gerontology. Series A, Biological sciences and medical sciences, 79(4), glad275. https://doi.org/10.1093/gerona/glad275 Phelan, J. C., Link, B. G., & Tehranifar, P. (2010). Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. Journal of health and social behavior, 51 Suppl, S28–S40. doi: 10.1177/0022146510383498. PMID: 20943581. Ramiller, A. (2024, April 1). Introduction to propensity score matching with MatchIt. D-Lab. https://dlab.berkeley.edu/news/introduction-propensity-score-matching-matchit Rewak, M., Buka, S., Prescott, J., De Vivo, I., Loucks, E. B., Kawachi, I., Non, A. L., & Kubzansky, L. D. (2014). Race-related health disparities and biological aging: does rate of telomere shortening differ across blacks and whites?. Biological psychology, 99, 92–99. Romaniuk, A., Paszel-Jaworska, A., Totoń, E., Lisiak, N., Hołysz, H., Królak, A., Grodecka-Gazdecka, S., & Rubiś, B. (2019). The non-canonical functions of telomerase: to turn off or not to turn off. Molecular biology reports, 46(1), 1401–1411. https://doi.org/10.1007/s11033-018-4496-x Selvaraju, V., Phillips, M., Fouty, A., Babu, J. R., & Geetha, T. (2021, January 9). Telomere length as a biomarker for race-related health disparities. MDPI. https://www.mdpi.com/2073-4425/12/1/78 Shammas, M. A. (2011). Telomeres, lifestyle, cancer, and aging. Current opinion in clinical nutrition and metabolic care. doi: 10.1097/MCO.0b013e32834121b1. PMID: 21102320; PMCID: PMC3370421. Shin, D., Shin, J., & Lee, K. W. (2019). Effects of Inflammation and Depression on Telomere Length in Young Adults in the United States. Journal of clinical medicine, 8(5), 711. doi: 10.3390/jcm8050711. PMID: 31109116; PMCID: PMC6572156. Solorio, S., Murillo-Ortíz, B., Hernández-González, M., Guillén-Contreras, J., Arenas-Aranda, D., Solorzano-Zepeda, F. J., Ruiz-Avila, R., Mora-Villalpando, C., de la Roca-Chiapas, J. M., & MalacaraHernández, J. M. (2011). Association between telomere length and C-reactive protein and the development of coronary collateral circulation in patients with coronary artery disease. Angiology, 62(6), 467–472. doi: 10.1177/0003319710398007. PMID: 21441231. Wang, F., Giskeødegård, G. F., Skarra, S., Engstrøm, M. J., Hagen, L., Geisler, J., Mikkola, T. S., Tikkanen, M. J., Debik, J., Reidunsdatter, R. J., & Bathen, T. F. (2023). Association of serum cortisol and cortisone levels and risk of recurrence after endocrine treatment in breast cancer. Clinical and experimental medicine, 23(7), 3883–3893. https://doi.org/10.1007/s10238-023-01109-x Wang, Z., Zhang, Z., Guo, Y., Shui, H., Liu, G., Jin, T., & Wang, H. (2018). Shorter Telomere Length Is Associated with Increased Breast Cancer Risk in a Chinese Han Population: A Case-Control Analysis. Journal of breast cancer, 21(4), 391–398. https://doi.org/10.4048/jbc.2018.21.e52 Williams, D. R., Priest, N., & Anderson, N. B. (2016). Understanding associations among race, socioeconomic status, and health: Patterns and prospects. Health psychology : official journal of the Division of Health Psychology, American Psychological Association, 35(4), 407–411. doi: 10.1037/hea0000242. PMID: 27018733; PMCID: PMC4817358. Wong, J. Y., De Vivo, I., Lin, X., Fang, S. C., & Christiani, D. C. (2014). The relationship between inflammatory biomarkers and telomere length in an occupational prospective cohort study. PloS one, 9(1), e87348. doi: 10.1371/journal.pone.0087348. PMID: 24475279; PMCID: PMC3903646. Yedjou, C. G., Sims, J. N., Miele, L., Noubissi, F., Lowe, L., Fonseca, D. D., Alo, R. A., Payton, M., & Tchounwou, P. B. (2019). Health and Racial Disparity in Breast Cancer. Advances in Experimental Medicine and Biology, 1152, 31. https://doi.org/10.1007/978-3-030-20301-6_3 Zheng, N.S., Annis, J., Master, H. et al. Sleep patterns and risk of chronic disease as measured by long-term monitoring with commercial wearable devices in the All of Us Research Program. Nat Med 30, 2648–2656 (2024). https://doi.org/10.1038/s41591-024-03155-8