Socioeconomic disparities in head and neck cancer survival in Germany: A causal mediation analysis using population-based cancer registry data.

Bedir, Ahmed¹; Abera, Semaw Ferede¹; Efremov, Ljupcho¹,²; Hassan, Lamiaa²; Vordermark, Dirk¹,³; Medenwald, Daniel¹,³

1. Department of Radiation Oncology, Health Services Research Group, University Hospital Halle (Saale), Ernst-Grube-Str. 40, 06120, Halle (Saale), Germany.
2. Institute of Medical Epidemiology, Biometry, and Informatics, Martin Luther University Halle-Wittenberg, Magdeburger Strasse 8, 06112, Halle (Saale), Germany.
3. Department of Radiation Oncology, University Hospital Halle (Saale), Ernst-Grube-Str. 40, 06120, Halle (Saale), Germany.

Address for correspondence:
Daniel Medenwald
Department of Radiation Oncology, University Hospital Halle (Saale), Ernst-Grube-Str. 40, 06120, Halle (Saale), Germany.
Telephone no: +49-345-557-3453/4027
Email: Daniel.Medenwald@uk-halle.de
Appendix 3: R Code for Mediation Analysis.

1. We first fitted a linear regression for the first mediator (m1) (medical care “loghosbeds”) and a logistic regression for the third mediator (m3) (treatment “tx”) conditional on the exposure (socioeconomic deprivation level “SD”), potential earlier intermediates (Stage at diagnosis “stage_dx”), and our confounders (age, sex, and year of diagnosis).

\[
\text{fitm1} \leftarrow \text{glm}(\text{loghosbeds} \sim \text{factor(SD) + age + sex + diagyear}, \text{family} = \text{gaussian("identity")}, \text{data} = \text{hanm1})
\]

\[
\text{fitm3} \leftarrow \text{glm}((\text{tx} \sim \text{factor(SD) + loghosbeds + factor(stage_dx)) + age + sex + diagyear, \text{family} = \text{binomial}, \text{data} = \text{hanm1})}
\]

2. We then fitted a logistic regression for the outcome variable (death at 6 months, 1 year conditional on 6-month survival, 2 years conditional on 1-year survival, or 5 years conditional on 2-year survival “TOD”).

\[
\text{fitY} \leftarrow \text{glm}(\text{TOD} \sim \text{factor(SD) + loghosbeds + factor(stage_dx) + factor(tx) + age + sex + diagyear, \text{family} = \text{binomial}, \text{data} = \text{hanm1})}
\]

3. We then extended our original dataset 8 times. Considering that the exposure levels analyzed at any one time were 2 (most affluent vs any other level) and we had 3 mediators \(2^3 = 8\). The extended dataset also included hypothetical exposure levels: a0, a1, and a2.

\[
\text{expData} \leftarrow \text{data.frame(replicate = rep(1:8, times = nrow(hanm1)), hanm1[rep(hanm1$id, each = 8), ], a0=NA, a1=NA, a2=NA, a3=NA)}
\]

\[
\text{expData} \leftarrow \text{within(expData, \{}
\]

\[
a2 \leftarrow \text{A}
\]

\[
a0 \leftarrow \text{ifelse(replicate \%in\% c(2,4,6,8), 1-A, A)}
\]

\[
a1 \leftarrow \text{ifelse(replicate \%in\% c(3,4,7,8), 1-A, A)}
\]

\[
a3 \leftarrow \text{ifelse(replicate \%in\% c(5,6,7,8), 1-A, A)}
\]

\[
\text{\})}
\]

4. We then calculated regression weights.

\[
\text{num1} \leftarrow \text{with(expData, dnorm(loghosbeds, mean = predict(fitm1, newdata = within(expData, A \leftarrow a1), type = "response"), sd = sqrt(summary(fitm1)$dispersion)))}
\]

\[
\text{num2} \leftarrow \text{with(expData, predict(fitm3, newdata = within(expData, A \leftarrow a3), type = "response"))}
\]

\[
\text{denom1} \leftarrow \text{with(expData, dnorm(loghosbeds, mean = predict(fitm1, newdata = within(expData, A \leftarrow a2), type = "response"), sd = sqrt(summary(fitm1)$dispersion)))}
\]

\[
\text{denom2} \leftarrow \text{with(expData, predict(fitm3, newdata = within(expData, A \leftarrow a2), type = "response"))}
\]

\[
\text{expData$W1} \leftarrow \text{num1*num2)/(denom1*denom2)}
\]
5. To calculation population-average component effects, we update the previously calculated weights by inverse weighting.

```r
fitA <- glm(SD~age+sex+diagyear, family = binomial("logit"), data=hanm1)

expData<-within (expData, {W1<-W1 / dbinom(A, size = 1, prob = predict(fitA, newdata=expData, type="response"))})
```

6. We finally fit a population-average natural effect model.

```r
fitNEMpop<- glm(TOD ~ a0*a1*a2*a3 , family=binomial("logit"), data = expData, weights = W1)
```

The analysis was repeated for each of the previously mentioned time points and only two quintiles were compared at a time.