

The effect of age on enthesal changes: a study of modifications at appendicular attachment sites in a large sample of identified human skeletons

Supporting Information (A)

Sébastien Villotte, Frédéric Santos*

October 20, 2022

Contents

1	Configuration	2
1.1	Details about the R session	2
1.2	Custom R functions	3
2	Load data	6
3	General approach for each enthesal locus	7
3.1	Mosaic plots and barplots	7
3.2	Relationship between age and enthesal changes	7
3.3	Ordinal logistic regression models	9
3.4	Main effect of age	9
4	Detailed application for the loci HDE_L and HEL_L	11
4.1	Descriptive statistics	11
4.2	Plot for severe enthesal changes (stage C)	12
4.3	Descriptive statistics (continued)	13
4.4	Main effect of sex (for locus HDE_L)	14
4.4.1	Ordinal logistic regression	14
4.4.2	Log-linear models	14
4.5	Main effect of age (for locus HDE_L)	16
4.5.1	Male individuals	16
4.5.2	Female individuals	16
5	Supplementary results	16
5.1	χ^2 residuals	16

*sebastien.villotte@cnrs.fr, frederic.santos@u-bordeaux.fr

1. Configuration

1.1. Details about the R session

This document aims to facilitate the replication of a subset of results (tables or figures) presented in our article. As stated in the main text of the article, our data are hosted on Zenodo (Villotte, 2022). All the analyses were performed using R (R Core Team, 2022), and this document has been built with Org mode 9.5.5 for Emacs 28.2 (Schulte, Davison, Dye, & Dominik, 2012).

Along with R version 4.2.1 (2022-06-23) itself, the following R packages are loaded, using their version available on CRAN at a fixed date (2022-10-01), thanks to the `{groundhog}` R package (Simonsohn & Gruson, 2021) :

```
## Use groundhog to improve the reproducibility:
library(groundhog)
gday <- "2022-10-01"
## Load the following packages:
groundhog.library("car", date = gday)
groundhog.library("gdata", date = gday)
groundhog.library("genodds", date = gday)
groundhog.library("here", date = gday)
groundhog.library("MASS", date = gday)
groundhog.library("tidyverse", date = gday)
groundhog.library("vcd", date = gday)
```

Additional details about the R session:

```
print(sessionInfo(), locale = FALSE)
```

```
R version 4.2.1 (2022-06-23)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Manjaro Linux

Matrix products: default
BLAS: /usr/lib/libopenblas-r0.3.21.so
LAPACK: /usr/lib/liblapack.so.3.10.1

attached base packages:
[1] grid      stats      graphics  grDevices  utils      datasets  methods
[8] base

other attached packages:
 [1] vcd_1.4-10      forcats_0.5.2  stringr_1.4.1  dplyr_1.0.10
 [5] purrr_0.3.4    readr_2.1.2    tidyr_1.2.1    tibble_3.1.8
 [9] ggplot2_3.3.6  tidyverse_1.3.2 MASS_7.3-58.1  here_1.0.1
[13] genodds_1.1.2  gdata_2.18.0.1 car_3.1-0      carData_3.0-5
[17] groundhog_2.1.0
```

1.2. Custom R functions

Several additional R functions were written. The first one allows for the comparison of the enthesal changes across age groups (using Fisher exact tests and generalised odds ratios), and produces a summary similar to Tables 7, 8, 10, 11 in the main text of our article.

```
compare_ag <- function(data, site = "HDE_L") {
  ## 1. Initialize an empty dataframe of results:
  res <- combn(x = levels(data$Age_group), m = 2) |>
    t() |>
    as.data.frame() |>
    rename(Group1 = V1, Group2 = V2) |>
    mutate(p_value_Fisher = NA) |>
    mutate(Gen_OR = NA)
  ## 2. Perform Fisher tests & Compute generalized OR:
  for (i in 1:nrow(res)) {
    temp <- subset(data, Age_group %in% res[i, 1:2]) |>
      droplevels()
    if (nlevels(temp[, site]) < 2) {
      return(NULL)
    }
    res[i, 3] <- round(
      fisher.test(temp$Age_group, temp[, site])$p.value,
      digits = 3
    )
    genor <- genodds(
      response = temp[, site],
      group = temp$Age_group
    )
    res[i, 4] <- paste0(
      "GenOR = ",
      round(exp(genor$pooled_lnodds), 3),
      "; IC = [",
      round(exp(genor$pooled_lnconf.int)[1], 3),
      ", ",
      round(exp(genor$pooled_lnconf.int)[2], 3),
      "]; p = ",
      round(genor$pooled_p, 3)
    )
  }
  ## Return dataframe of results:
  return(res)
}
```

The second one, `eta2()`, allows for the computation of correlation ratios (usually denoted η^2) between a continuous and a categorical variable.

```
var_inter <- function(x, groups) {
  ### x: numeric vector; groups: a factor
  moyennes <- tapply(x, INDEX = groups, FUN = mean)
  effectifs <- tapply(x, INDEX = groups, FUN = length)
  return((sum(effectifs * (moyennes - mean(x))^2)) / length(x))
}

var_tot <- function(x) {
  return(sum((x - mean(x))^2) / length(x))
}

eta2 <- function(x, groups) {
  ### x: numeric vector; groups: a factor
  ### Computes the correlation ratio between x and groups
  dtf <- data.frame(x = x, groups = groups)
  if (any(is.na(dtf))) {
    dtf <- na.omit(dtf)
  }
  return(var_inter(dtf$x, dtf$groups) / var_tot(dtf$x))
}
```

The R code below allows for the generation of barplots that describe, for each sex, the repartition of age classes, in a similar spirit of Figure 9 in Listi (2016). This is used for Figure 3 from our main text. Note that this graphical representation only makes sense because all age groups have a roughly similar number of individuals in the whole sample.

```
### Barplots for severe enthesal changes (Fig. 3 in main text):
cp2 <- function(x) {return(sum(x == "C", na.rm = TRUE))}
palette(c("navy", "royalblue", "yellow", "orange", "red"))
barplot2 <- function(data, sex, legend = TRUE, sites = c("HDE_L", "HEL_L")) {
  temp <- subset(data, Sex == sex)
  tab <- apply(
    X = temp[, sites],
    MARGIN = 2,
    FUN = function(x) tapply(x, FUN = cp2, INDEX = temp$Age_group)
  ) |>
  prop.table(margin = 2)
  barplot(100 * tab, beside = FALSE, col = 1:5,
    ylab = "Percent",
    legend.text = legend,
    main = paste0("Frequencies of stage C by age group (", sex, ")"))
}
```

Finally, the R code below displays a table summarizing Pearson residuals and standardized residuals after a χ^2 test, in a similar spirit of Table 1 from (Sharpe, 2015).

```
chisq_resid <- function(age, locus) {  
  ### age, locus: two factors.  
  tab <- table(age, locus)  
  chi <- chisq.test(tab)  
  res <- chi$residuals  
  std <- chi$stdres  
  ## Put all results in the same table:  
  int <- interleave(tab, res, std) |>  
    as.matrix() |>  
    round(2)  
  intd <- data.frame(Age = rownames(int),  
                    Type = c("Observed", "Pearson res.", "Std. res."),  
                    int, row.names = NULL)  
  return(intd)  
}
```

2. Load data

We first load data into R, and summarize the dataframe.

```
## Load the CSV file:
dat <- read.csv2(
  file = "https://zenodo.org/record/6805623/files/data.csv?download=1",
  na.strings = "",
  row.names = 1,
  stringsAsFactors = TRUE
) |>
  mutate(across(HSC_R:TSO_L, as.factor))
## Summarize the dataset:
summary(dat)
```

Collection	Sex	Age	Age_group	HSC_R	HSI_R
Bologne	: 84 F:248	Min. :20.00	[20-29]:168	A :373	A :416
Coimbra	:326 M:473	1st Qu.:30.00	[30-39]:142	B :207	B :140
Sardaigne	:133	Median :44.00	[40-49]:128	C : 33	C : 17
Spitalfields	:178	Mean :44.93	[50-59]:130	NA's:108	NA's:148
		3rd Qu.:57.00	[60-99]:153		
		Max. :91.00			

HEL_R	HEM_R	RBB_R	HSC_L	HSI_L	HEL_L	HEM_L
A :359	A :440	A :437	A :413	A :451	A :417	A :434
B :138	B : 67	B :172	B :180	B :102	B :113	B : 59
C : 47	C : 8	C : 20	C : 17	C : 12	C : 21	C : 5
NA's:177	NA's:206	NA's: 92	NA's:111	NA's:156	NA's:170	NA's:223

RBB_L	CSB_R	FPF_R	FMF_R	FIP_R	CSB_L	FPF_L
A :436	A :482	A :548	A :483	A :430	A :481	A :541
B :160	B :130	B : 77	B :101	B :160	B :141	B : 70
C : 35	C : 39	C : 21	C : 7	C : 31	C : 24	C : 26
NA's: 90	NA's: 70	NA's: 75	NA's:130	NA's:100	NA's: 75	NA's: 84

FMF_L	FIP_L	HGP_R	HDE_R	RRP_R	HGP_L	HDE_L
A :501	A :420	A :124	A :250	A :265	A :149	A :244
B : 92	B :152	B :344	B :286	B :284	B :338	B :289
C : 11	C : 18	C :100	C : 55	C : 18	C : 52	C : 53
NA's:117	NA's:131	NA's:153	NA's:130	NA's:154	NA's:182	NA's:135

RRP_L	FGF_R	FLA_R	TSO_R	FGF_L	FLA_L	TSO_L
A :266	A :223	A :244	A :392	A :223	A :235	A :371
B :278	B :342	B :354	B :230	B :331	B :339	B :222
C : 24	C : 78	C : 47	C : 37	C : 81	C : 55	C : 51
NA's:153	NA's: 78	NA's: 76	NA's: 62	NA's: 86	NA's: 92	NA's: 77

We replicate here the Table 2 from our main text (description of the population samples):

```
## Replication of table 2 from the main text:
with(dat, ftable(Collection, Sex, Age_group))
```

		Age_group	[20-29]	[30-39]	[40-49]	[50-59]	[60-99]
Collection	Sex						
Bologne	F		0	0	0	0	0
	M		20	16	13	18	17
Coimbra	F		38	31	25	28	31
	M		48	50	33	24	18
Saraigne	F		0	0	0	0	0
	M		45	19	31	21	17
Spitalfields	F		11	10	14	22	38
	M		6	16	12	17	32

Finally, we convert all enthesal loci into *ordered* factors:

```
## Convert enthesal sites into ordered factors:
dat <- mutate(dat, across(HSC_R:TSO_L, factor, ordered = TRUE))
```

3. General approach for each enthesal locus

3.1. Mosaic plots and barplots

Barplots (e.g., Figure 1 of our main text) and mosaic plots (Friendly, 1994) (e.g., Figure 2 of our main text) are first created to visualize the distribution of enthesal changes across age groups, and their relationship with sex. See Section 4.1 of the present document for an example of R code for the case of the enthesis HDE_L.

3.2. Relationship between age and enthesal changes

Table 3 below summarizes part of Tables 3–4 from our main text, by giving the Spearman correlation coefficient between age-at-death and the scores of enthesal changes (as an ordinal variable). Alternatively, a correlation ratio, denoted η^2 , is also given for the relationship between age-at-death and enthesal changes (as a nominal variable) (Saporta, 2011).

```
## Initialize an empty matrix:
sites <- colnames(dat)[-c(1:4)]
matcor <- matrix(NA, nrow = length(sites), ncol = 2)
rownames(matcor) <- sites
colnames(matcor) <- c("Sperman coef.", "eta2")

## Fill in the matrix:
for (j in sites) {
  matcor[j, 1] <- cor(x = as.numeric(dat$Age),
                     y = as.numeric(dat[, j]),
                     method = "spearman",
                     use = "complete.obs")
}
```

```

    matcor[j, 2] <- eta2(x = dat$Age, groups = dat[, j])
}

## Display the resulting matrix:
print(round(matcor, 3))

```

	Sperman coef.	eta2
HSC_R	0.506	0.267
HSI_R	0.442	0.214
HEL_R	0.628	0.401
HEM_R	0.415	0.197
RBB_R	0.404	0.181
HSC_L	0.476	0.238
HSI_L	0.466	0.246
HEL_L	0.583	0.377
HEM_L	0.405	0.19
RBB_L	0.461	0.227
CSB_R	0.464	0.232
FPF_R	0.431	0.21
FMF_R	0.471	0.242
FIP_R	0.421	0.206
CSB_L	0.476	0.242
FPF_L	0.427	0.204
FMF_L	0.42	0.192
FIP_L	0.439	0.206
HGP_R	0.509	0.261
HDE_R	0.498	0.238
RRP_R	0.528	0.255
HGP_L	0.517	0.254
HDE_L	0.48	0.216
RRP_L	0.526	0.256
FGF_R	0.64	0.381
FLA_R	0.64	0.391
TSO_R	0.372	0.132
FGF_L	0.649	0.395
FLA_L	0.629	0.373
TSO_L	0.391	0.149

Table 3: Spearman coefficients and correlation ratios (η^2) between age (or age groups) and enthesal changes at each locus.

3.3. Ordinal logistic regression models

For each enthesal site, an ordinal logistic regression model is defined as follows:

$$\text{site} \sim \text{Age} + \text{Sex} + \text{Age}:\text{Sex}$$

In Table 5 (in our main text and below), we summarize the results of such models for each enthesal site, by testing the main effects of the two covariates (age-at-death and sex) along with their interaction term.

```
### Replication of Table 5 from the main text:
## Initialize an empty matrix:
resolr <- matrix(NA, ncol = 3, nrow = length(sites))
rownames(resolr) <- sites
colnames(resolr) <- c("Age", "Sex", "Age x Sex")
## Build OLR models:
for (s in sites) {
  olrs <- polr(
    formula = as.formula(paste(s, "~ Age * Sex")),
    data = dat
  )
  resolr[s, ] <- Anova(olrs)$"Pr(>Chisq)"
}
## Print table of results:
round(print(resolr), 3)
```

3.4. Main effect of age

The main effect of age is already proved *via* the ordinal logistic regression models above. However, it can be assessed and described more precisely. We adopted two different approaches:

1. A Fisher exact test is performed to compare each pair of age groups (e.g., [20-29] vs. [30-39], etc.), thus offering some kind of *post-hoc* tests after having proved the global effect of age — a process known as *ransacking* (Sharpe, 2015).
2. Furthermore, for each pair of age groups, a generalized odds ratio is computed (Agresti, 1980) along with its 95% confidence interval; see main text for a detailed explanation.

	Age	Sex	Age x Sex
HSC_R	0	0.003	0.001
HSI_R	0	0.986	0.191
HEL_R	0	0	0.266
HEM_R	0	0.004	0.388
RBB_R	0	0	0.456
HSC_L	0	0	0.742
HSI_L	0	0.764	0.325
HEL_L	0	0.001	0.779
HEM_L	0	0.038	0.162
RBB_L	0	0	0.086
CSB_R	0	0.042	0.542
FPF_R	0	0.007	0.491
FMF_R	0	0.261	0.372
FIP_R	0	0	0.358
CSB_L	0	0.058	0.802
FPF_L	0	0.001	0.885
FMF_L	0	0.944	0.373
FIP_L	0	0	0.693
HGP_R	0	0	0.396
HDE_R	0	0	0.272
RRP_R	0	0	0.438
HGP_L	0	0	0.995
HDE_L	0	0	0.435
RRP_L	0	0	0.855
FGF_R	0	0.084	0.043
FLA_R	0	0	0.019
TSO_R	0	0.001	0.19
FGF_L	0	0.529	0.143
FLA_L	0	0	0.018
TSO_L	0	0.006	0.929

Table 5: p -values (rounded at the third decimal place) for the main effects of Age and Sex, and for the interaction between Age and Sex, in an ordinal logistic regression model with the corresponding enthesal site as response variable.

4. Detailed application for the loci HDE_L and HEL_L

4.1. Descriptive statistics

A stacked barplot describing the distribution of enthesal stages across age groups in the whole sample can be found in Figure 1 (below and in our main text).

```
### Replication of Figure 1:  
layout(mat = matrix(c(1, 2), ncol = 2), widths = c(0.45, 0.55))  
with(dat, table(HDE_L, Age_group)) |>  
  prop.table(margin = 2) |>  
  apply(MARGIN = 2, FUN = "*", 100) |>  
  subset(c(FALSE, TRUE, TRUE)) |>  
  barplot(col = c("cornflowerblue", "navy"), ylim = c(0, 100),  
          ylab = "% of individuals", xlab = "Age group",  
          main = "Distribution of enthesal changes for HDE_L")  
with(dat, table(HEL_L, Age_group)) |>  
  prop.table(margin = 2) |>  
  apply(MARGIN = 2, FUN = "*", 100) |>  
  subset(c(FALSE, TRUE, TRUE)) |>  
  barplot(col = c("cornflowerblue", "navy"),  
          legend.text = TRUE,  
          args.legend = list(x = 7, y = 95),  
          xlim = c(0, 7), ylim = c(0, 100),  
          ylab = "% of individuals", xlab = "Age group",  
          main = "Distribution of enthesal changes for HEL_L")
```

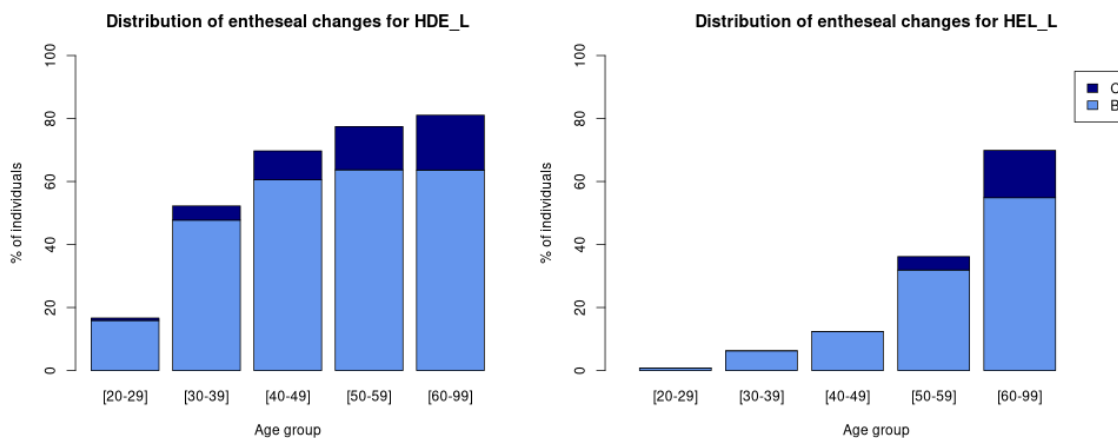


Figure 1: Stacked barplot for the distribution of enthesal changes across age groups.

More precisely, a *mosaic plot* describing the distribution of stages of HDE_L by age and sex can be found in Figure 2 (below and in our main text).

```
## Replication of Figure 2:
mosaic(
  ~ Sex + Age_group + HDE_L,
  data = dat,
  highlighting = "HDE_L",
  highlighting_fill = c("lightblue", "cornflowerblue", "navy")
)
```

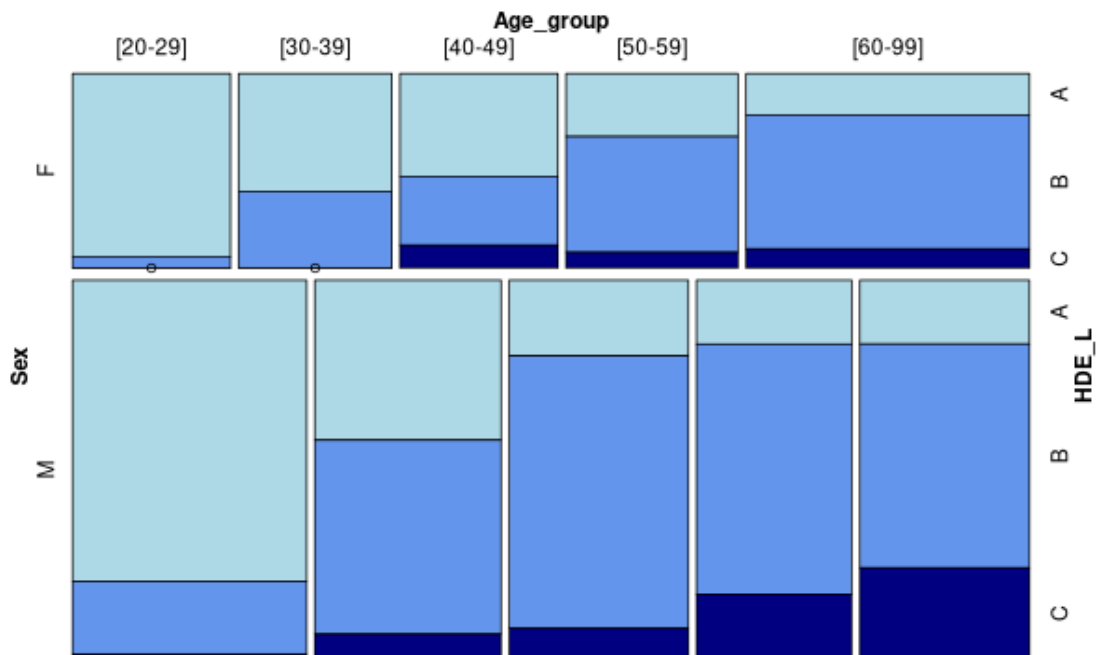


Figure 2: Mosaic plot for enthesal changes in HDE_L locus by sex and age group.

4.2. Plot for severe enthesal changes (stage C)

The R code below allows for the replication of Figure 3 in our main text.

```
## Replication of Figure 3:
par(mfrow = c(1, 2))
barplot2(dat, sex = "F", legend = FALSE)
barplot2(dat, sex = "M")
```

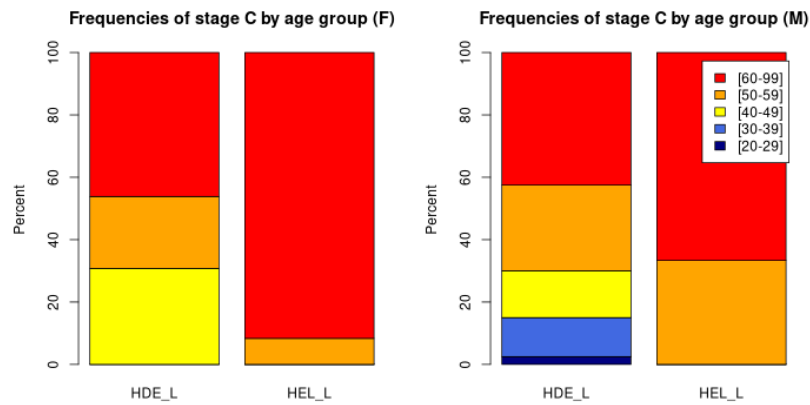


Figure 3: Relative frequencies of the most severe enthesal modification (stage C) by age group, depending on sex.

4.3. Descriptive statistics (continued)

The R code below allows for the replication of Figure 4 for our main text.

```
## Replication of Figure 4:
mosaic(
  ~ Sex + Age_group + HEL_L,
  data = dat,
  highlighting = "HEL_L",
  highlighting_fill = c("lightblue", "cornflowerblue", "navy")
)
```

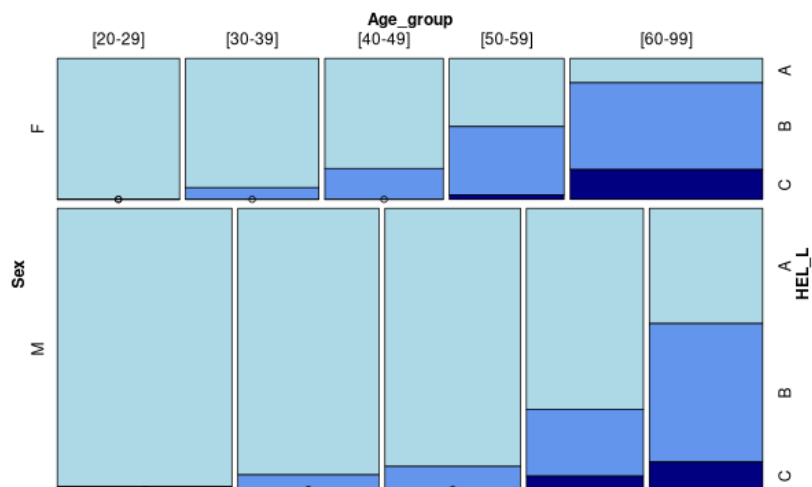


Figure 4: Mosaic plot for enthesal changes in HEL_L locus by sex and age group.

4.4. Main effect of sex (for locus HDE_L)

The main effect of sex can be assessed either by a logistic regression model or a log-linear model (Christensen, 1997).

4.4.1. Ordinal logistic regression

The analysis of deviance table is given below:

```
## Logistic regression:
olr_delt <- polr(
  formula = HDE_L ~ Age_group * Sex,
  data = dat
)
Anova(olr_delt)
```

Analysis of Deviance Table (Type II tests)

Response: HDE_L

	LR	Chisq	Df	Pr(>Chisq)
Age_group	176.871	4	< 2.2e-16	***
Sex	21.523	1	3.497e-06	***
Age_group:Sex	1.351	4	0.8527	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The global conclusions would remain unchanged when using raw age (in years) instead of age groups, as illustrated below:

```
## Logistic regression:
olr_delt2 <- polr(
  formula = HDE_L ~ Age * Sex,
  data = dat
)
Anova(olr_delt2)
```

Analysis of Deviance Table (Type II tests)

Response: HDE_L

	LR	Chisq	Df	Pr(>Chisq)
Age	156.784	1	< 2.2e-16	***
Sex	26.683	1	2.397e-07	***
Age:Sex	0.609	1	0.4353	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.4.2. Log-linear models

Below, we first define a *saturated* log-linear model (i.e., including all possible main effects and interaction terms), and then we apply a stepwise procedure for model selection in order to find the best model according to this procedure.

```

## Turn data into a table of counts:
counts <- dat |>
  select(Sex, Age_group, HDE_L) |>
  ftable() |>
  as.data.frame()
## Saturated log-linear model:
lls_delt <- glm(Freq ~ Sex * Age_group * HDE_L,
               family = poisson, data = counts)
## Stepwise (backward) model selection:
llbest_delt <- step(lls_delt, direction = "backward")

```

```

Start:  AIC=185.52
Freq ~ Sex * Age_group * HDE_L

              Df Deviance   AIC
- Sex:Age_group:HDE_L  8   12.132 181.66
<none>                    0.000 185.52

Step:  AIC=181.65
Freq ~ Sex + Age_group + HDE_L + Sex:Age_group + Sex:HDE_L +
      Age_group:HDE_L

              Df Deviance   AIC
<none>                    12.132 181.65
- Sex:HDE_L                2   33.115 198.64
- Sex:Age_group            4   41.577 203.10
- Age_group:HDE_L         8  188.464 341.99

```

This suggests that the three-way interaction term (Sex:Age_group:HDE_L) may not be included in the model, i.e. that sex and age do not interact relative to enthesal changes on the locus, a result which is consistent with the logistic regression model above.

Alternatively, one can also show that this interaction term is not relevant using a test between two nested log-linear models, as illustrated below:

```

## Submodel without three-way interaction:
llr_delt <- update(lls_delt, ~ . - (Sex:Age_group:HDE_L))
## Test between the two nested models:
anova(lls_delt, llr_delt, test = "Chi")

```

Analysis of Deviance Table

```

Model 1: Freq ~ Sex * Age_group * HDE_L
Model 2: Freq ~ Sex + Age_group + HDE_L + Sex:Age_group + Sex:HDE_L +
      Age_group:HDE_L
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         0     0.000
2         8    12.132 -8  -12.132   0.1454

```

4.5. Main effect of age (for locus HDE_L)

4.5.1. Male individuals

Table 6 (in our main text and below) presents the same results for female individuals only.

```
## Compare age groups (males only):  
compare_ag(data = subset(dat, Sex == "M"), site = "HDE_L")
```

	Group1	Group2	p_value_Fisher	Gen_OR
1	[20-29]	[30-39]	0	GenOR = 2.226; IC = [1.62, 3.059]; p = 0
2	[20-29]	[40-49]	0	GenOR = 4.052; IC = [2.768, 5.93]; p = 0
3	[20-29]	[50-59]	0	GenOR = 4.759; IC = [3.139, 7.218]; p = 0
4	[20-29]	[60-99]	0	GenOR = 5.011; IC = [3.317, 7.571]; p = 0
5	[30-39]	[40-49]	0.011	GenOR = 1.557; IC = [1.13, 2.147]; p = 0.007
6	[30-39]	[50-59]	0.002	GenOR = 1.85; IC = [1.316, 2.6]; p = 0
7	[30-39]	[60-99]	0	GenOR = 2.025; IC = [1.438, 2.853]; p = 0
8	[40-49]	[50-59]	0.273	GenOR = 1.221; IC = [0.892, 1.673]; p = 0.212
9	[40-49]	[60-99]	0.03	GenOR = 1.37; IC = [0.995, 1.886]; p = 0.054
10	[50-59]	[60-99]	0.59	GenOR = 1.124; IC = [0.805, 1.571]; p = 0.492

Table 6: Pairwise comparisons of age groups using Fisher exact tests (uncorrected p -values) and Generalized odds ratios, for males only.

4.5.2. Female individuals

Table 7 (in our main text and below) presents the same results for female individuals only.

```
## Compare age groups (females only):  
compare_ag(data = subset(dat, Sex == "F"), site = "HDE_L")
```

5. Supplementary results

The following results are not presented in our main text, and give further insight about the relationship between age and enthesal changes.

5.1. χ^2 residuals

Following Sharpe (2015), Tables 8 and 9 provide the χ^2 residuals for the association between age groups and enthesal stages.

```
## Chi-squared residuals for HDE_L:  
chisq_resid(age = dat$Age_group, locus = dat$HDE_L)
```


	Group1	Group2	p_value_Fisher	Gen_OR
1	[20-29]	[30-39]	0.001	GenOR = 2.008; IC = [1.325, 3.043]; p = 0.001
2	[20-29]	[40-49]	0	GenOR = 2.44; IC = [1.562, 3.814]; p = 0
3	[20-29]	[50-59]	0	GenOR = 4.286; IC = [2.475, 7.422]; p = 0
4	[20-29]	[60-99]	0	GenOR = 6.514; IC = [3.777, 11.236]; p = 0
5	[30-39]	[40-49]	0.165	GenOR = 1.28; IC = [0.788, 2.081]; p = 0.318
6	[30-39]	[50-59]	0.023	GenOR = 1.914; IC = [1.168, 3.137]; p = 0.01
7	[30-39]	[60-99]	0	GenOR = 2.519; IC = [1.59, 3.992]; p = 0
8	[40-49]	[50-59]	0.135	GenOR = 1.392; IC = [0.843, 2.298]; p = 0.197
9	[40-49]	[60-99]	0.003	GenOR = 1.74; IC = [1.07, 2.828]; p = 0.025
10	[50-59]	[60-99]	0.498	GenOR = 1.257; IC = [0.837, 1.888]; p = 0.27

Table 7: Pairwise comparisons of age groups using Fisher exact tests (uncorrected p -values) and Generalized odds ratios, for females only.

	Age	Type	A	B	C
1	[20-29]	Observed	110	21	1
2	[20-29]	Pearson res.	7.42	-5.47	-3.17
3	[20-29]	Std. res.	11.04	-8.72	-3.77
4	[30-39]	Observed	53	53	5
5	[30-39]	Pearson res.	1	-0.24	-1.59
6	[30-39]	Std. res.	1.45	-0.37	-1.85
7	[40-49]	Observed	33	66	10
8	[40-49]	Pearson res.	-1.84	1.67	0.05
9	[40-49]	Std. res.	-2.67	2.6	0.05
10	[50-59]	Observed	23	65	14
11	[50-59]	Pearson res.	-2.99	2.07	1.57
12	[50-59]	Std. res.	-4.3	3.2	1.81
13	[60-99]	Observed	25	84	23
14	[60-99]	Pearson res.	-4.04	2.34	3.2
15	[60-99]	Std. res.	-6.01	3.74	3.81

Table 8: Left deltoid tuberosity: contingency table for stages and age groups. ("Observed:" observed frequency; "Pearson res.": Pearson residuals in χ^2 test; "Std. Res.": standardized residuals in χ^2 test.)

```
## Chi-squared residuals for HEL_L:
chisq_resid(age = dat$Age_group, locus = dat$HEL_L)
```

	Age	Type	A	B	C
1	[20-29]	Observed	126	1	0
2	[20-29]	Pearson res.	3.05	-4.91	-2.2
3	[20-29]	Std. res.	7.05	-6.27	-2.56
4	[30-39]	Observed	105	7	0
5	[30-39]	Pearson res.	2.2	-3.33	-2.07
6	[30-39]	Std. res.	4.99	-4.19	-2.36
7	[40-49]	Observed	92	13	0
8	[40-49]	Pearson res.	1.41	-1.84	-2
9	[40-49]	Std. res.	3.17	-2.29	-2.27
10	[50-59]	Observed	60	30	4
11	[50-59]	Pearson res.	-1.32	2.44	0.22
12	[50-59]	Std. res.	-2.94	3.01	0.25
13	[60-99]	Observed	34	62	17
14	[60-99]	Pearson res.	-5.57	8.07	6.12
15	[60-99]	Std. res.	-12.67	10.15	6.99

Table 9: Left deltoid tuberosity: contingency table for stages and age groups. (“Observed:” observed frequency; “Pearson res.”: Pearson residuals in χ^2 test; “Std. Res.”: standardized residuals in χ^2 test.)

References

- Agresti, A. (1980). Generalized Odds Ratios for Ordinal Data. *Biometrics*, 36(1), 59. doi:10.2307/2530495
- Christensen, R. (1997). *Log-linear models and logistic regression* (Second ed.). New York: Springer.
- Friendly, M. (1994). Mosaic Displays for Multi-Way Contingency Tables. *Journal of the American Statistical Association*, 89(425), 190–200. doi:10.1080/01621459.1994.10476460
- Listi, G. A. (2016). The Use of Entheseal Changes in the Femur and Os Coxa for Age Assessment. *Journal of Forensic Sciences*, 61(1), 12–18. doi:10.1111/1556-4029.12905
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Saporta, G. (2011). *Probabilités, analyse des données et statistique* (2e ed.). Paris: Technip.
- Schulte, E., Davison, D., Dye, T., & Dominik, C. (2012). A Multi-Language Computing Environment for Literate Programming and Reproducible Research. *Journal of Statistical Software*, 46(1), 1–24. doi:10.18637/jss.v046.i03
- Sharpe, D. (2015). Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research, and Evaluation*, 20(1). doi:10.7275/tbfa-x148
- Simonsohn, U., & Gruson, H. (2021). *Groundhog: Reproducible Scripts via Version-Specific Package Loading*. Retrieved from <https://CRAN.R-project.org/package=groundhog>
- Villotte, S. (2022). *Data set EC Age*. Zenodo. doi:10.5281/ZENODO.6805622