

Patients with immune-mediated inflammatory diseases receiving cytokine inhibitors have low prevalence of SARS-CoV-2 seroconversion

Source data and analysis code to supplement submission to Nature Communications.

See author list below

July 3, 2020

David Simon, Koray Tascilar, Gerhard Krönke, Arnd Kleyer, Mario M. Zaiss, Franz Heppt, Christine Meder, Raja Atreya, Entcho Klenske, Peter Dietrich, Abdullah Abdullah, Thorsten Kliem, Giulia Corte, Harriet Morf, Moritz Leppkes, Andreas E. Kremer, Andreas Ramming, Milena Pachowsky, Florian Schuch, Monika Ronneberger, Stefan Kleinert, Clara Maier, Axel J. Hueber, Karin Manger, Bernhard Manger, Carola Berking, Matthias Tenbusch, Klaus Überla, Michael Sticherling, Markus F. Neurath, and Georg Schett

1 Load packages, read data, define custom functions

```
library(tidyverse)
library(readxl)
library(epitools)
library(sandwich)
library(kableExtra)
library(ggbeeswarm)
library(janitor)
library(tables)
library(broom)
library(epitools)

booktabs(titlerule="\\hline")

firstdata<- read_excel("sarscov2seroprevalencedata.xlsx",
                      guess_max = 2000)
#pcrresults <- read_excel("pcrresults.xlsx")
#socialdistancing <- read_excel("socialdistancing.xlsx") %>% distinct() %>%
# group_by(tscID) %>%
# filter(SocialDistancing%in% c("ja", "nein"))

#positivecontrols <- read_excel("positivecontrols.xlsx")
cumcase <- read_excel("erlangencumulativecases.xlsx") %>%
  clean_names() %>%
  mutate(total=erlangen_stadt+hochstadt) %>%
  select(datum, totalcumcase=total)

coviddatax <- firstdata %>%
  rename(oldtscID=tscID) %>%
  mutate(tscID=1:n(),
         group=fct_relevel(fct_recode(group, `Healthcare Control`="Healthcare provider",
                                       `Non-Healthcare Control`="General population",
```

```

        `IMIDs Cytokine INH`="b-tsDMARD user",
        `IMIDs Non-Cytokine INH`="no b-tsDMARD user"),
c("Non-Healthcare Control",
  "Healthcare Control",
  "IMIDs Cytokine INH",
  "IMIDs Non-Cytokine INH"))) %>%

arrange(group)

theme_set(theme_classic())

```

```

formateffectsize <- function(point,lower,upper,rounding=2,trans=FALSE){
  if (trans) {
    point <- exp(point)
    lower <- exp(lower)
    upper <- exp(upper)
  }
  paste0(
    format(round(point,rounding),nsmall = rounding,trim = TRUE),
    " (",
    format(round(lower,rounding),nsmall = rounding,trim = TRUE),
    " to ",
    format(round(upper,rounding),nsmall = rounding,trim = TRUE),
    ")",
    sep = ""
  )
}

```

```

pvalformat <- function(x,rounding=3){
  require(dplyr)
  case_when(
    x>0.99 ~ ">0.99",
    x<0.001 ~ "<0.001",
    TRUE ~ paste(format(round(x,rounding),nsmall = rounding))
  )
}

```

```

mymeansd <- function(x,rounding=1){
  a <- mean(x,na.rm = T)
  b <- sd(x,na.rm = T)
  if(is.na(a)|is.infinite(a)){
    paste("-")
  }else{
    paste(format(round(a,rounding),nsmall = rounding),
          " (",
          format(round(b,rounding),nsmall = rounding),
          ")",sep = "")
  }
}

```

```

npercent <- function(x,y,rounding=1){
  percentage=round(100*length(x)/length(y),rounding)
  if(is.na(percentage)|is.infinite(percentage)){
    paste("-")
  }else{
    paste(length(x),
          " (",

```

```
format(percentage, nsmall = rounding),
")",
sep = "")}}}
```

2 Description of the study group.

2.1 Baseline characteristics.

```
table1 <- coviddatav %>%
  rename(smoking=46, txcategory=57) %>%
  mutate(chronische_Atemwegserkr=
    as.numeric(chronische_Atemwegserkr%in%c(1,2)),
    dx=fct_lump_min(dx,min=20),
    txcategory=fct_lump_min(txcategory,min = 39)) %>%
  tabular(
    (N=1)*(` `=1)+

    (Age=age)*(` `=1)*(`mean(SD)`=mymeansd)+

    (Factor(sex,levelnames = c("Male","Female")))+

    Factor(smoking,levelnames = c("Never","Current","Past","Missing"))+

    (Diabetes=Factor(Diabetes_mellitus,levelnames = c("No","Yes","Missing")))+

    Factor(Hypertonie, levelnames = c("No","Yes","Missing"))+

    (`Pulmonary disease`=Factor(chronische_Atemwegserkr,levelnames = c("No","Yes")))+

    (Treatment=Factor(txcategory))+

    (Diagnosis=Factor(dx))*(`N(\\%)`=Percent(denom = Equal(group),fn = npercent))+

    (BMI*mymeansd)

    ~

    Factor(group),
    data=.

  )

write.csv.tabular(table1,"table1.csv")

#Write source data for table
coviddatav %>%
  rename(smoking=46, txcategory=57) %>%

  mutate(chronische_Atemwegserkr=
    as.numeric(chronische_Atemwegserkr%in%c(1,2)),
    dx=fct_lump_min(dx,min=20),
    txcategory=fct_lump_min(txcategory,min = 39)) %>%
  mutate(sex=factor(sex,labels = c("Male","Female"))) %>%
```

```

select(tscID,group, Age=age,sex,BMI,smoking,Diabetes=Diabetes_mellitus,
      Hypertension=Hypertonie,`Pulmonary disease`=chronische_Atemwegserkr,
      Treatment=txcategory,Diagnosis=dx) %>%
mutate_at(6:9,function(x) ifelse(x %in% c("keine Angabe","999"),NA,x)) %>%

arrange(tscID) %>%
write.csv("table1sourcedata.csv")

```

Table 1: Baseline characteristics

			group			
4-7			Non-Healthcare Control	Healthcare Control	IMiDs Cytokine INH	IMiDs Non-Cytokine INH
	N		971	285	534	259
Age	mean(SD)		43.2 (14.3)	40.3 (12.7)	48.9 (15.7)	55.3 (16.1)
sex	Male	N(%)	697 (71.8)	94 (33.0)	248 (46.4)	105 (40.5)
	Female	N(%)	274 (28.2)	189 (66.3)	285 (53.4)	152 (58.7)
smoking	Never	N(%)	596 (61.4)	178 (62.5)	317 (59.4)	105 (40.5)
	Current	N(%)	181 (18.6)	35 (12.3)	94 (17.6)	40 (15.4)
	Past	N(%)	125 (12.9)	0 (0.0)	42 (7.9)	41 (15.8)
	Missing	N(%)	8 (0.8)	0 (0.0)	17 (3.2)	9 (3.5)
Diabetes	No	N(%)	854 (88.0)	200 (70.2)	450 (84.3)	211 (81.5)
	Yes	N(%)	59 (6.1)	12 (4.2)	42 (7.9)	14 (5.4)
	Missing	N(%)	0 (0.0)	0 (0.0)	1 (0.2)	0 (0.0)
Hypertonie	No	N(%)	796 (82.0)	204 (71.6)	349 (65.4)	150 (57.9)
	Yes	N(%)	117 (12.0)	8 (2.8)	145 (27.2)	75 (29.0)
	Missing	N(%)	0 (0.0)	0 (0.0)	1 (0.2)	0 (0.0)
Pulmonary disease	No	N(%)	904 (93.1)	278 (97.5)	488 (91.4)	243 (93.8)
	Yes	N(%)	67 (6.9)	7 (2.5)	46 (8.6)	16 (6.2)
Treatment	IL-17 Inhibitors	N(%)	0 (0.0)	0 (0.0)	51 (9.6)	0 (0.0)
	IL-23 Inhibitors	N(%)	0 (0.0)	0 (0.0)	85 (15.9)	0 (0.0)
	IL-6 Inhibitors	N(%)	0 (0.0)	0 (0.0)	44 (8.2)	0 (0.0)
	JAK Inhibitors	N(%)	0 (0.0)	0 (0.0)	39 (7.3)	0 (0.0)
	Others	N(%)	0 (0.0)	0 (0.0)	88 (16.5)	0 (0.0)
	TNF Inhibitors	N(%)	0 (0.0)	0 (0.0)	227 (42.5)	0 (0.0)
Diagnosis	IBD	N(%)	0 (0.0)	0 (0.0)	176 (33.0)	14 (5.4)
	Psoriasis	N(%)	0 (0.0)	0 (0.0)	63 (11.8)	28 (10.8)
	RA	N(%)	0 (0.0)	0 (0.0)	130 (24.3)	106 (40.9)
	SpA	N(%)	0 (0.0)	0 (0.0)	117 (21.9)	34 (13.1)
	Other	N(%)	0 (0.0)	0 (0.0)	48 (9.0)	77 (29.7)
	BMI	mymeansd	26.5 (6.0)	23.6 (4.4)	26.4 (5.8)	26.4 (4.5)

2.2 Seropositivity stratified by group, age, sex and other covariates

```

agecattable <- coviddatax %>%
  mutate(serostatus=factor(value>=0.8,
                           labels = c("IgG (-)", "IgG(+)")),
         agecat=cut2(floor(age), cuts = c(20,40,60,80))) %>%
  tabular(((Heading(Age)*Factor(agecat,
                              levelnames = c("<=19", "20-39",
                                             "40-59", "60-79", ">=80"),
                              name = "Age"))+
          (Sex=Factor(sex, levelnames = c("Male", "Female")))+
          (Diabetes=Factor(Diabetes_mellitus,
                          levelnames = c("No", "Yes", "Not reported")))+
          (Hypertension=Factor(Hypertonie,
                              levelnames = c("No", "Yes", "Not reported")))+
          (`Pulmonary disease`=Factor(chronische_Atemwegserkr,
                                      levelnames = c("No", "Yes", "Not reported"))))
  ) *
  Heading()*serostatus
~
(Heading()*Percent(denom = Equal(group, agecat, Diabetes_mellitus, Hypertonie, chronische_Atemwegserkr, sex), f
Heading()*group, data = .)

#Save table as csv file
write.csv.tabular(agecattable, "seropositivity_breakdown.csv")

#Save source data as csv file
coviddatax %>%
  mutate(serostatus=factor(value>=0.8,
                           labels = c("IgG (-)", "IgG(+)")),
         agecat=cut2(floor(age), cuts = c(20,40,60,80)),
         agecat=factor(agecat, labels = c("<=19", "20-39",
                                         "40-59", "60-79", ">=80")),
         sex=factor(sex, labels = c("Male", "Female")),
         Diabetes_mellitus=factor(Diabetes_mellitus, labels = c("No", "Yes", "Not reported")),
         Hypertonie=factor(Hypertonie, labels = c("No", "Yes", "Not reported")),
         chronische_Atemwegserkr=factor(chronische_Atemwegserkr, labels = c("No", "Yes", "Not reported"))) %>%
  select(tscID, group, agecat, sex,
         Diabetes_mellitus, Hypertension=Hypertonie,
         `Chronic lung disease`=chronische_Atemwegserkr,
         OD450nm=value, serostatus) %>%
  arrange(tscID) %>%
  write.csv("table2sourcedata.csv", row.names = F)

```

Table 2: Seropositivity by characteristics

		Non-Healthcare Control		Healthcare Control	IMIDs Cytokine INH	IMIDs Non-Cytokine INH
Age	<=19	IgG (-)	47 (97.9)	1 (100.0)	5 (100.0)	2 (100.0)
		IgG(+)	1 (2.1)	0 (0.0)	0 (0.0)	0 (0.0)
	20-39	IgG (-)	334 (97.9)	141 (97.9)	152 (98.7)	44 (97.8)
		IgG(+)	7 (2.1)	3 (2.1)	2 (1.3)	1 (2.2)
	40-59	IgG (-)	461 (97.9)	109 (94.0)	228 (100.0)	93 (96.9)
		IgG(+)	10 (2.1)	7 (6.0)	0 (0.0)	3 (3.1)
	60-79	IgG (-)	101 (97.1)	22 (91.7)	136 (98.6)	99 (98.0)
		IgG(+)	3 (2.9)	2 (8.3)	2 (1.4)	2 (2.0)
	>=80	IgG (-)	6 (85.7)	-	9 (100.0)	13 (86.7)
		IgG(+)	1 (14.3)	-	0 (0.0)	2 (13.3)
Sex	Male	IgG (-)	681 (97.7)	87 (92.6)	245 (98.8)	102 (97.1)
		IgG(+)	16 (2.3)	7 (7.4)	3 (1.2)	3 (2.9)
	Female	IgG (-)	268 (97.8)	184 (97.4)	284 (99.6)	147 (96.7)
		IgG(+)	6 (2.2)	5 (2.6)	1 (0.4)	5 (3.3)
Diabetes	No	IgG (-)	837 (98.0)	191 (95.5)	447 (99.3)	204 (96.7)
		IgG(+)	17 (2.0)	9 (4.5)	3 (0.7)	7 (3.3)
	Yes	IgG (-)	56 (94.9)	11 (91.7)	41 (97.6)	14 (100.0)
		IgG(+)	3 (5.1)	1 (8.3)	1 (2.4)	0 (0.0)
	Not reported	IgG (-)	-	-	1 (100.0)	-
		IgG(+)	-	-	0 (0.0)	-
Hypertension	No	IgG (-)	780 (98.0)	194 (95.1)	347 (99.4)	144 (96.0)
		IgG(+)	16 (2.0)	10 (4.9)	2 (0.6)	6 (4.0)
	Yes	IgG (-)	113 (96.6)	8 (100.0)	143 (98.6)	74 (98.7)
		IgG(+)	4 (3.4)	0 (0.0)	2 (1.4)	1 (1.3)
	Not reported	IgG (-)	-	-	1 (100.0)	-
		IgG(+)	-	-	0 (0.0)	-
Pulmonary disease	No	IgG (-)	829 (98.0)	195 (95.6)	440 (99.1)	203 (97.1)
		IgG(+)	17 (2.0)	9 (4.4)	4 (0.9)	6 (2.9)
	Yes	IgG (-)	64 (95.5)	6 (85.7)	46 (100.0)	15 (93.8)
		IgG(+)	3 (4.5)	1 (14.3)	0 (0.0)	1 (6.2)
	Not reported	IgG (-)	-	-	1 (100.0)	-
		IgG(+)	-	-	0 (0.0)	-

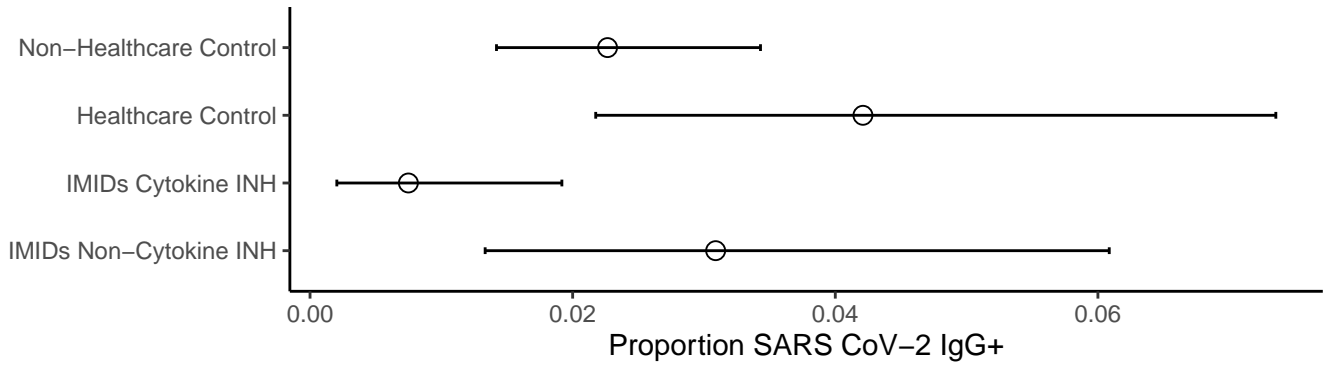


Figure 1: Crude SARS-CoV-2 seroprevalance and confidence intervals.

Table 3: Crude seroprevalence of SARS-CoV-2 and Poisson approximated 95% confidence intervals

Group	N	Positive	Prevalence (95% CI)
Non-Healthcare Control	971	22	0.0227 (0.0142 to 0.0343)
Healthcare Control	285	12	0.0421 (0.0218 to 0.0735)
IMIDs Cytokine INH	534	4	0.0075 (0.0020 to 0.0192)
IMIDs Non-Cytokine INH	259	8	0.0309 (0.0133 to 0.0609)

3 Calculate prevalence and 95% confidence intervals

```

coviddatax %>%
  split(.$group) %>%
  map(~pois.exact(x = sum(.$value>=0.8),pt = length(.$value))) %>%
  bind_rows(.id = "group") %>%
  mutate(group=fct_relevel(group,
                           c(levels(coviddatax$group))
  )) %>%

  ggplot(aes(fct_rev(group),rate)) +
  geom_point(shape=1, size=3)+
  geom_errorbar(aes(ymin=lower,ymax=upper),width=0.1)+
  labs(y="Proportion SARS CoV-2 IgG+",x=NULL)+
  coord_flip()

```

```

coviddatax %>%
  mutate(persontime=1) %>%
  split(.$group) %>%
  map(~epitools::pois.exact(x = sum(.$value>=0.8),pt = length(.$value))) %>%
  bind_rows(.id = "group") %>%
  mutate(prevalence=format(effects::rate,lower,upper,rounding = 4)) %>%
  select(Group=group, `N`=pt, `Positive`=x, `Prevalence (95% CI)`=prevalence) %>%
  kable(booktabs=T,
        caption = "Crude seroprevalence of SARS-CoV-2 and
        Poisson approximated 95\\% confidence intervals") %>%
  kable_styling()

```

4 Estimate adjusted relative risk of SARS-CoV-2 seropositivity

Groups were compared using Poisson regression with robust standard errors to calculate age and sex adjusted risk ratios. The reference level for the comparison was the general population. This model adjusts for age, sex and the time of

Table 4: Risk ratios from Poisson model with robust variances. Adjusted by age, sex and cumulative case counts

Term	RR (95% CI)	P value
Healthcare Control	2.36 (1.03 to 5.43)	0.043
IMIDs Cytokine INH	0.32 (0.11 to 0.99)	0.048
IMIDs Non-Cytokine INH	1.21 (0.50 to 2.90)	0.676
Age, 10 year intervals	1.23 (0.98 to 1.56)	0.080
Female sex	0.63 (0.33 to 1.18)	0.149
Cumulative case count	1.00 (1.00 to 1.00)	>0.99

sampling. Instead of the number of days from the earliest sampling date, we used the total cumulative number of cases in Erlangen/Erlangen-Höchststadt on the date of sampling for adjustment, this would reflect the probability of being exposed to SARS-CoV2.

```
#Fit model
m08 <- coviddat %>%
  left_join(cumcase, by=c("testdate"="datum")) %>%
  mutate(sex=factor(sex, labels = c("Male", "Female"))) %>%

  glm((value>=0.8)~group +
      I((age-mean(age))/10)+
      sex+totalcumcase, family = "poisson", data = .)

#Estimate heteroskedasticity consistent covariance matrix
cov.m08 <- vcovHC(m08, type="HC0")

#Estimate sandwich standard errors
std.err <- sqrt(diag(cov.m08))

#Estimate coefficients, 95% confidence intervals,
#p values and exponentiate to get risk ratios.
r.est08 <- tibble(Term=names(coef(m08)),
  Estimate= coef(m08), "Robust SE" = std.err,
  "Pr(>|z|)" = 2 * pnorm(abs(coef(m08)/std.err),
    lower.tail=FALSE),
  LL = coef(m08) - 1.96 * std.err,
  UL = coef(m08) + 1.96 * std.err) %>%
  mutate(result=formatEFFECTSIZE(Estimate, LL, UL, trans = T),
    p=pvalformat(`Pr(>|z|)`))

r.est08 %>%
  select(Term, `RR (95% CI)`=result, `P value`=p) %>%
  slice(-1) %>%
  mutate(Term=c("Healthcare Control",
    "IMIDs Cytokine INH",
    "IMIDs Non-Cytokine INH",
    "Age, 10 year intervals",
    "Female sex",
    "Cumulative case count")) %>%
  kable(booktabs=T, caption = "Risk ratios from Poisson model with robust
    variances. Adjusted by age, sex and cumulative case counts") %>%
  kable_styling()
```

5 Supplementary figure 1, symptoms in seropositive participants clinically diagnosed vs. not diagnosed with COVID-19

```
coviddatax %>%
  select(tscID,value,clinicalcovid,29:32,34:38,40,41) %>%
  pivot_longer(-c(1:3), names_to = "symptom",values_to = "symptomstat") %>%
  filter(value>=0.8) %>%
  mutate(tscID=factor(tscID)) %>%
  mutate(symptomstat=case_when(
    symptomstat%in%c("1","ja")~"1",
    symptomstat%in%c("0","nein","keine Angabe")~"0",
    TRUE~"Not Assessed"
  )) %>%
  group_by(tscID) %>%
  mutate(nsymptom=sum(as.numeric(symptomstat),na.rm = T)) %>%
  ungroup() %>%
  mutate(tscID=fct_reorder(factor(tscID),nsymptom,.fun =)) %>%
  ggplot(aes(symptom,factor(tscID)))+
  geom_tile(aes(fill=symptomstat),colour="gray90") +
  facet_grid(fct_rev(factor(clinicalcovid,labels = c("Undiagnosed","Diagnosed"))))~.,
            scales = "free",
            space = "free_y",switch = "both") +
  theme(axis.line.y = element_blank(),
        axis.ticks.y = element_blank(),
        axis.title.x = element_blank(),
        axis.text.y = element_blank())+
  labs(fill=NULL,y=NULL)+
  scale_x_discrete(position = "top",guide = guide_axis(n.dodge = 2)) +
  scale_fill_manual(values = c("#3E2D96", "#DDE637", "#ABABAB"))

## Warning: NAs introduced by coercion
## Warning: NAs introduced by coercion
## Warning: NAs introduced by coercion
## Warning: NAs introduced by coercion
## Warning: NAs introduced by coercion
```

5.1 Plot deviations from expected symptom frequencies

This plots standardized Pearson's residuals from the chisquared test. For the "significant" tests, this plot gives an idea about which groups cause the significance. The labels can be fixed if we want to use it.

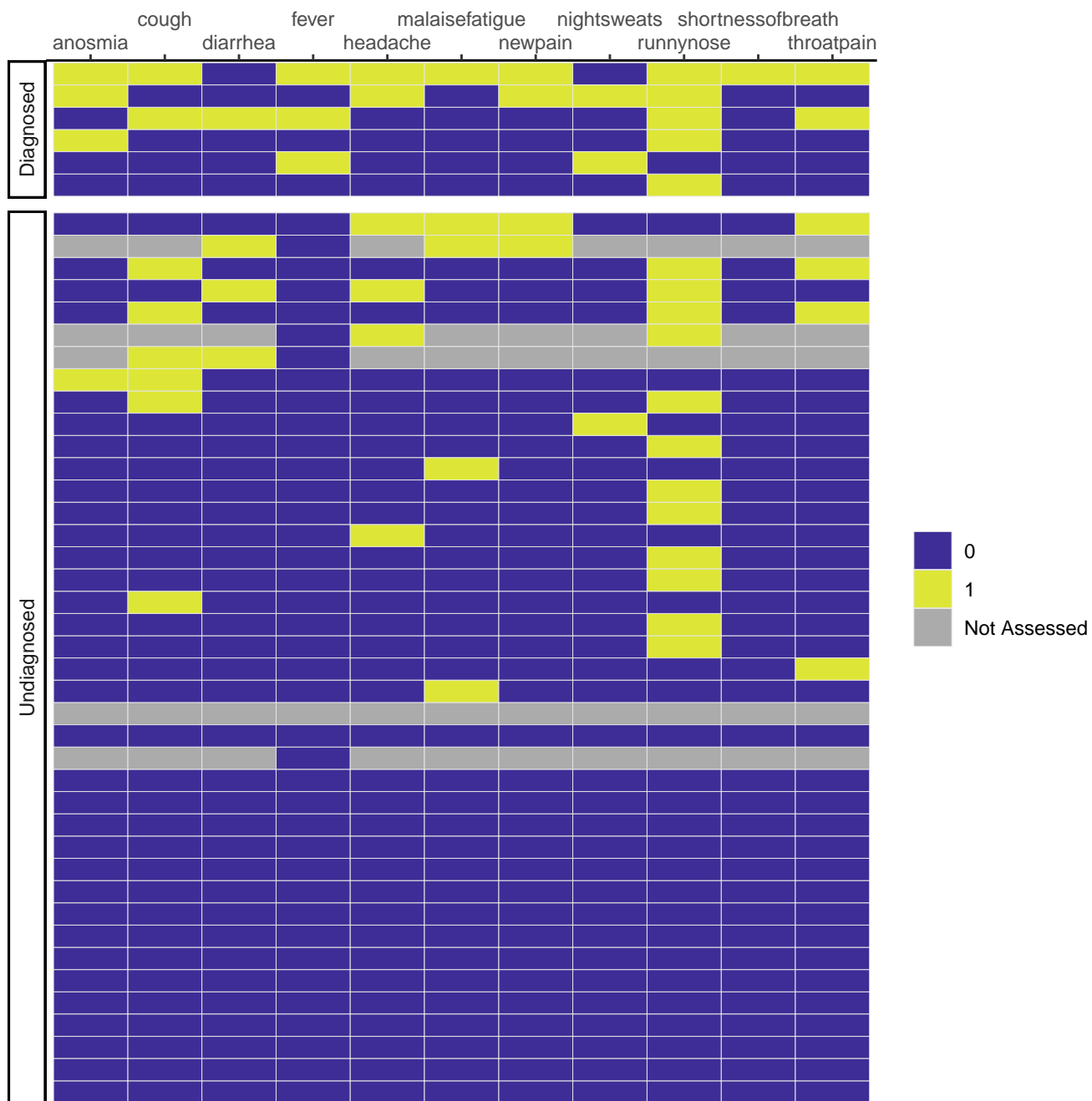


Figure 2: Symptoms of the seropositive

6 Estimate standardized residuals for exposure risk behaviour.

```
#Preparedata for pearson residuals plot
dodgewidth <- 0.7
pivotcols <- c(27,28,64)
coviddatatax %>%
  pivot_longer(cols = pivotcols,
               names_to = "variable",
               values_to = "status_pre",
               values_ptypes = list(status_pre = 'character')) %>%
  filter(!(status_pre %in% c("No data", "keine Angabe", NA))) %>%
  mutate(status=case_when(status_pre %in% c("1", "ja", "Yes",
                                             "Retired/Unemployed")~1,
                           TRUE~0),
         group=fct_rev(fct_relevel(group, c("IMID b-tsDMARD",
                                             "IMID n-csDMARD")))) %>%

group_by(variable) %>%

#Estimate residuals for each variable
do(with(., augment(chisq.test(table(group, status)))) %>%

#Get 0 values for HomeOffice, this indicates going to work
filter(status==ifelse(variable=="lhoHomeOffice", 0, 1)) %>%
ungroup() %>%

mutate(variable=factor(variable,
                       labels = c("Contact", "Workplace",
                                   "Risk area")),
       group=fct_rev(fct_relevel(group, c("IMIDs Cytokine INH",
                                           "IMIDs Non-Cytokine INH",
                                           "Non-Healthcare Control",
                                           "Healthcare Control")))) %>%

#Plot residuals
ggplot(aes(1, .stdres, group=group))+
  geom_point( position = position_dodge(width = dodgewidth))+
  geom_hline(yintercept = 0)+
  geom_linerange(aes(ymin=0, ymax=.stdres),
                position = position_dodge(width = dodgewidth))+
  coord_flip()+
  scale_shape(solid = T)+
  facet_wrap(variable~., ncol = 3, strip.position = "top")+
  scale_y_continuous(breaks=c(-5, 0, 5, 10, 15), labels = c(-5, 0, 5, 10, 15))+
```

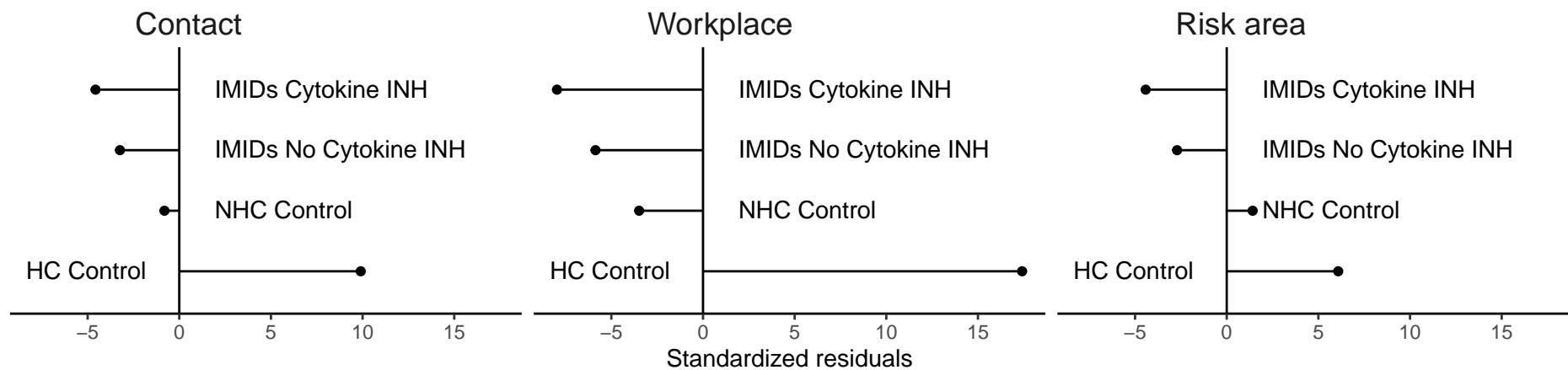


Figure 3: Standardized residuals showing deviation of contact with febrile respiratory infections, presence in workplace and travel to risk areas and contact with documented COVID 19 cases from expected frequencies for each group.

```
theme(strip.background = element_blank(),
      axis.text.y = element_blank(),
      axis.line.y = element_blank(),
      axis.ticks.y = element_blank(),
      strip.text = element_text(hjust = 0.3, size = 14)) +

annotate("text", x=0.74, y=-1.8,
         label="HC Control", hjust=1, vjust=0.5) +
annotate("text", x=0.915, y=1.94,
         label="NHC Control", hjust=0, vjust=0.5) +
annotate("text", x=1.09, y=1.94,
         label="IMIDs No Cytokine INH", hjust=0, vjust=0.5) +
annotate("text", x=1.265, y=1.94,
         label="IMIDs Cytokine INH", hjust=0, vjust=0.5) +

labs(y="Standardized residuals", x=NULL, shape=NULL)

## Note: Using an external vector in selections is ambiguous.
## i Use 'all_of(pivotcols)' instead of 'pivotcols' to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
## Warning: Unknown levels in 'f': IMID b-tsDMARD, IMID n-csDMARD
```

7 Breakdown of symptoms for groups and IgG status

```
tablefinegroups<- coviddatax %>%
  mutate(serostatus=factor(value>=0.8,
                            labels = c("IgG (-)", "IgG(+)", "")) %>%
  tabular((N=1)*(` `=1)*(` `=1)+
    (`New MSK pain`=Factor(newpain%in%c("1", "ja"),
                           levelnames =c("No", "Yes")))*
    Percent(denom = Equal(group), fn = npercent)+

    (`Night sweats`=Factor(nightsweats%in%c("1", "ja"),
                           levelnames =c("No", "Yes")))*
    Percent(denom = Equal(group), fn = npercent)+

    (`Fever`=Factor( fever%in%c("1", "ja"),
                    levelnames =c("No", "Yes")))*
    Percent(denom = Equal(group), fn = npercent)+

    (`Malaise/Fatigue`=Factor(malaisefatigue%in%c("1", "ja"),
                              levelnames =c("No", "Yes")))*
    Percent(denom = Equal(group), fn = npercent)+

    (`Headache`=Factor(headache%in%c("1", "ja"),
                       levelnames =c("No", "Yes")))*
    Percent(denom = Equal(group), fn = npercent)+

    (`Rhinitis`=Factor(runny nose%in%c("1", "ja"),
                      levelnames =c("No", "Yes")))*
    Percent(denom = Equal(group), fn = npercent)+

    (`Shortness of breath`=Factor(shortnessofbreath%in%c("1", "ja"),
                                  levelnames =c("No", "Yes")))*
    Percent(denom = Equal(group), fn = npercent)+

    (`Cough`=Factor(cough%in%c("1", "ja"),
                    levelnames =c("No", "Yes")))*
    Percent(denom = Equal(group), fn = npercent)+

    (`Throat pain`=Factor(throatpain%in%c("1", "ja"),
                          levelnames =c("No", "Yes")))*
    Percent(denom = Equal(group), fn = npercent)+

    (`Anosmia`=Factor( anosmia%in%c("1", "ja"),
```

```

        levelnames =c("No", "Yes")))*
Percent(denom = Equal(group),fn = npercent)+

(`Diarrhea`=Factor(diarrhea%in%c("1", "ja"),
        levelnames =c("No", "Yes")))*
Percent(denom = Equal(group),fn = npercent)

~

Heading()*Factor(group),
data = . )

latex(tablefinegroups,options =
list(tabular="longtable",
toprule="\caption{Symptoms by serostatus and group}\\\\\\\\\\toprule",
midrule="\midrule\\\\\\[-1\\normalbaselineskip]\\endhead\\hline\\endfoot"))

```

Table 5: Symptoms by serostatus and group

			Non-Healthcare Control	Healthcare Control	IMIDs Cytokine INH	IMIDs Non-Cytokine INH
N			971	285	534	259
New MSk pain	No	Percent	903 (93.0)	266 (93.3)	477 (89.3)	228 (88.0)
	Yes	Percent	68 (7.0)	19 (6.7)	57 (10.7)	31 (12.0)
Night sweats	No	Percent	912 (93.9)	254 (89.1)	488 (91.4)	222 (85.7)
	Yes	Percent	59 (6.1)	31 (10.9)	46 (8.6)	37 (14.3)
Fever	No	Percent	913 (94.0)	270 (94.7)	508 (95.1)	244 (94.2)
	Yes	Percent	58 (6.0)	15 (5.3)	26 (4.9)	15 (5.8)
Malaise/Fatigue	No	Percent	877 (90.3)	217 (76.1)	447 (83.7)	223 (86.1)
	Yes	Percent	94 (9.7)	68 (23.9)	87 (16.3)	36 (13.9)
Headache	No	Percent	755 (77.8)	188 (66.0)	415 (77.7)	215 (83.0)
	Yes	Percent	216 (22.2)	97 (34.0)	119 (22.3)	44 (17.0)
Rhinitis	No	Percent	663 (68.3)	153 (53.7)	393 (73.6)	222 (85.7)
	Yes	Percent	308 (31.7)	132 (46.3)	141 (26.4)	37 (14.3)
Shortness of breath	No	Percent	919 (94.6)	269 (94.4)	494 (92.5)	236 (91.1)
	Yes	Percent	52 (5.4)	16 (5.6)	40 (7.5)	23 (8.9)
Cough	No	Percent	815 (83.9)	218 (76.5)	462 (86.5)	224 (86.5)
	Yes	Percent	156 (16.1)	67 (23.5)	72 (13.5)	35 (13.5)
Throat pain	No	Percent	756 (77.9)	195 (68.4)	445 (83.3)	231 (89.2)
	Yes	Percent	215 (22.1)	90 (31.6)	89 (16.7)	28 (10.8)
Anosmia	No	Percent	951 (97.9)	279 (97.9)	522 (97.8)	252 (97.3)
	Yes	Percent	20 (2.1)	6 (2.1)	12 (2.2)	7 (2.7)
Diarrhea	No	Percent	894 (92.1)	256 (89.8)	449 (84.1)	234 (90.3)
	Yes	Percent	77 (7.9)	29 (10.2)	85 (15.9)	25 (9.7)

```

write.csv.tabular(tablefinegroups,"symptombreakdown.csv")

#write source data to file

coviddatav %>%
  select(tscID,group,`New musculoskeletal pain`=newpain,
         `Night sweats`=nightsweats,Fever=fever,`Malaise/Fatigue`=malaisefatigue,
         Headache=headache, Rhinitis=runnynose,`Shortness of breath`=shortnessofbreath,
         Cough=cough,`Throat pain`=throatpain,Anosmia=anosmia,diarrhea=diarrhea) %>%
  mutate_at(3:13, function(x) case_when(
    x %in% c("1","ja") ~ "Yes",
    x %in% c("0","nein") ~ "No",
    TRUE ~ NA_character_
  )) %>% write.csv("symptomdata.csv")

```