

```

x<-read_excel(ofile2,sheet=osheet2,col_types="text")
x<-x[,c(1,3)]
colnames(x)=c("CAS","SMILES")
x$id <- as.character(1:nrow(x))

data1a<-read_excel(ofile1,sheet=osheet1,col_types="text")
blank<-length(data1a)+1
data1a[blank] <- NA
if("#" %in% colnames(data1a)){c9<-which(colnames(data1a)=="#")}
if("CAS Number" %in% colnames(data1a)){c1<-which(colnames(data1a)=="CAS
Number")}
if("Endpoint" %in% colnames(data1a)){c2<-
which(colnames(data1a)=="Endpoint")}
if("Reference source" %in% colnames(data1a)){c3<-
which(colnames(data1a)=="Reference source")}
if("Test organisms (species)" %in% colnames(data1a)){c4<-
which(colnames(data1a)=="Test organisms (species)")}
if("Assay" %in% colnames(data1a)){c5<-which(colnames(data1a)=="Assay")}
if("Toxicity Effect" %in% colnames(data1a)){c6<-
which(colnames(data1a)=="Toxicity Effect")}
if("Type of method" %in% colnames(data1a)){c7<-
which(colnames(data1a)=="Type of method")}
if("#" %notin% colnames(data1a)){c9<-blank}
if("CAS Number" %notin% colnames(data1a)){c1<-blank}
if("Endpoint" %notin% colnames(data1a)){c2<-blank}
if("Reference source" %notin% colnames(data1a)){c3<-blank}
if("Test organisms (species)" %notin% colnames(data1a)){c4<-blank}
if("Assay" %notin% colnames(data1a)){c5<-blank}
if("Toxicity Effect" %notin% colnames(data1a)){c6<-blank}
if("Type of method" %notin% colnames(data1a)){c7<-blank}
data<-cbind(data1a[,c(c1,c2,c3,c4,c5,c6,c7,c9)])
colnames(data)=c("CAS","End","Ref","Spec","Assay","TE","Type","id")
for(a in 1:dim(data)[1]){
  if(data$CAS[a]=="No CAS number"){data$CAS[a]<-""}}

d2<-inner_join(tlist,data[,-8],by="CAS")
d3<-filter(d2,d2$Spec %notin% c("Trout","Chicken") & d2$End != "AC50" &
d2$CAS != "")
if(dim(d3)[1]!=0){
  for (i in 1:dim(d3)[1]) {
    if(d3$End[i] %in% c('chemosis score','conjunctivae score','cornea
opacity score',
                        'iris score','overall irritation score','edema
score','erythema score',
                        'primary dermal irritation index
(PDII)','MMAS','Primary Irritation Index')){
      d3$animal[i]<-1
      d3$a_irr[i]<-1}
    if(d3$End[i]=='S W A N'){
      d3$animal[i]<-1
      d3$a_sens[i]<-1}
    if(d3$End[i] %in% c('Carcinogenicity','Summary
carcinogenicity','Summary Carcinogenicity')){
      d3$qual[i]<-1

```

```

    d3$ql_carc[i]<-1
    d3$animal[i]<-1
    d3$a_carc[i]<-1}
if(d3$End[i] %in% c('Eye Irritation','Skin Irritation/Corrosion')){
    d3$qual[i]<-1
    d3$ql_irr[i]<-1
    d3$animal[i]<-1
    d3$a_irr[i]<-1}
if(d3$End[i] %in% c('Chromosome aberration','Chromosome
Aberration','Gene mutation',
                    'In vitro cytogenicity / chromosome aberration
study in mammalian cells')){
    d3$qual[i]<-1
    d3$ql_muta[i]<-1
    d3$animal[i]<-1
    d3$a_muta[i]<-1}
if(d3$End[i]=='Developmental toxicity'){
    d3$qual[i]<-1
    d3$ql_repro[i]<-1
    d3$animal[i]<-1
    d3$a_repro[i]<-1}
if(d3$End[i]=='Skin sensitisation'){
    d3$qual[i]<-1
    d3$ql_sens[i]<-1
    d3$animal[i]<-1
    d3$a_sens[i]<-1}
if(d3$End[i]=='LC50'){
    d3$quan[i]<-1
    d3$qn_acute[i]<-1
    d3$animal[i]<-1
    d3$a_acute[i]<-1}
if(d3$End[i]=='Other Endpoint'){
    d3$quan[i]<-1
    d3$qn_syst[i]<-1
    d3$animal[i]<-1
    d3$a_syst[i]<-1}
if(d3$End[i]=='LD50'){
    d3$quan[i]<-1
    d3$qn_acute[i]<-1
    d3$animal[i]<-1
    d3$a_acute[i]<-1}
if(d3$End[i] %in% c('LOEL','NOEL','NOEL calculated','NOEL
Calculated','NOAEL','BMDL10')){
    d3$quan[i]<-1
    d3$qn_syst[i]<-1
    d3$animal[i]<-1
    d3$a_syst[i]<-1}
if(d3$End[i]=='TD50'){
    d3$quan[i]<-1
    d3$qn_carc[i]<-1
    d3$animal[i]<-1
    d3$a_carc[i]<-1}
if(d3$End[i] %in% c('DNA and protein damage','Relative ARBA','Relative
ERBA',

```

```

      'Aromatase inhibition potency (IC50)', 'Aromatase
Inhibition Potency (IC50)',
      'CD54', 'CD86', 'DNA damage and repair', 'Cell
transformation')){
  d3$exp[i]<-1
  d3$e_conn[i]<-1}
  if(d3$End[i]=='IC50'){d3$exp[i]<-1}
  if(d3$End[i]=='Dermal absorption'){
    d3$tox[i]<-1
    d3$t_vitro[i]<-1}
  if(d3$End[i] %in% c('Half-life (biotransformation)', 'Half life
(biotransformation)')){
    d3$tox[i]<-1
    if(d3$Spec[i] %in% c('Homo sapiens', 'Human')){d3$t_hmet[i]<-1}
    else{d3$t_amet[i]<-1}}
  if(d3$End[i] %in% c('Half-life (body)', 'Half-life (total body)')){
    d3$tox[i]<-1
    if(d3$Spec[i] %in% c('Homo sapiens', 'Human')){d3$t_habs[i]<-1}
    else{d3$t_aabs[i]<-1}}
  if(d3$End[i] %in% c('EC1.5', 'EC2', 'EC3')){
    if(d3$Type[i]=="in Vivo"){
      d3$animal[i]<-1
      d3$a_sens[i]<-1}
    else if(d3$Type[i]=="in Vitro"){
      d3$exp[i]<-1
      d3$e_conn[i]<-1}}
  if(d3$End[i]=="Reproductive toxicity"){
    d3$qual[i]<-1
    d3$ql_repro[i]<-1
    if(d3$TE[i]!="No relevant studies identified"){
      d3$animal[i]<-1
      d3$a_repro[i]<-1}}
  if(d3$Spec[i] %in% c('Peromyscus polionotus', 'Cattle', 'Sheep')){
    d3[i,2:ncolt]<-0
    d3$exp[i]<-1
  }
}
}

```

```

d3$c<-paste0(as.character(d3$CAS), as.character(d3$Ref))
d4<-d3[,c(2,5:ncolt,(ncolt+7))]

```

```

n1<- d4 %>% group_by(c, CAS) %>%
  summarise(across(where(is.numeric), ~ max(.x)))

```

```

n2<-n1[,-1]
ncas<- n2 %>% group_by(CAS) %>%
  summarise_all(sum)
if(dim(d3)[1] != 0){all1<-
merge(ncas,filter(tlist,!is.na(tlist$CAS)),all.y=TRUE,by="CAS")}}
if(dim(d3)[1] == 0){all1<-filter(tlist,!is.na(tlist$CAS))}

```

```

d2<-inner_join(tlist,x[,-1],by="SMILES")
d2<-inner_join(d2,data[,-1],by="id")

```

```

d3<-filter(d2,d2$Spec %notin% c("Trout","Chicken") & d2$End != "AC50" &
d2$id != "")
if(dim(d3)[1]!=0){
  for (i in 1:dim(d3)[1]) {
    if(d3$End[i] %in% c('chemosis score','conjunctivae score','cornea
opacity score',
                        'iris score','overall irritation score','edema
score','erythema score',
                        'primary dermal irritation index
(PDII)','MMAS','Primary Irritation Index')){
      d3$animal[i]<-1
      d3$a_irr[i]<-1}
    if(d3$End[i]=='S W A N'){
      d3$animal[i]<-1
      d3$a_sens[i]<-1}
    if(d3$End[i] %in% c('Carcinogenicity','Summary
carcinogenicity','Summary Carcinogenicity')){
      d3$qual[i]<-1
      d3$ql_carc[i]<-1
      d3$animal[i]<-1
      d3$a_carc[i]<-1}
    if(d3$End[i] %in% c('Eye Irritation','Skin Irritation/Corrosion')){
      d3$qual[i]<-1
      d3$ql_irr[i]<-1
      d3$animal[i]<-1
      d3$a_irr[i]<-1}
    if(d3$End[i] %in% c('Chromosome aberration','Chromosome
Aberration','Gene mutation',
                        'In vitro cytogenicity / chromosome aberration
study in mammalian cells')){
      d3$qual[i]<-1
      d3$ql_muta[i]<-1
      d3$animal[i]<-1
      d3$a_muta[i]<-1}
    if(d3$End[i]=='Developmental toxicity'){
      d3$qual[i]<-1
      d3$ql_repro[i]<-1
      d3$animal[i]<-1
      d3$a_repro[i]<-1}
    if(d3$End[i]=='Skin sensitisation'){
      d3$qual[i]<-1
      d3$ql_sens[i]<-1
      d3$animal[i]<-1
      d3$a_sens[i]<-1}
    if(d3$End[i]=='LC50'){
      d3$quan[i]<-1
      d3$qn_acute[i]<-1
      d3$animal[i]<-1
      d3$a_acute[i]<-1}
    if(d3$End[i]=='Other Endpoint'){
      d3$quan[i]<-1
      d3$qn_syst[i]<-1
      d3$animal[i]<-1
      d3$a_syst[i]<-1}
  }
}

```

```

    if(d3$End[i]=='LD50'){
      d3$quan[i]<-1
      d3$qn_acute[i]<-1
      d3$animal[i]<-1
      d3$a_acute[i]<-1}
    if(d3$End[i] %in% c('LOEL','NOEL','NOEL calculated','NOEL
Calculated','NOAEL','BMDL10')){
      d3$quan[i]<-1
      d3$qn_syst[i]<-1
      d3$animal[i]<-1
      d3$a_syst[i]<-1}
    if(d3$End[i]=='TD50'){
      d3$quan[i]<-1
      d3$qn_carc[i]<-1
      d3$animal[i]<-1
      d3$a_carc[i]<-1}
    if(d3$End[i] %in% c('DNA and protein damage','Relative
ARBA','Relative ERBA',
                        'Aromatase inhibition potency (IC50)','Aromatase
Inhibition Potency (IC50)',
                        'CD54','CD86','DNA damage and repair','Cell
transformation')){
      d3$exp[i]<-1
      d3$e_conn[i]<-1}
    if(d3$End[i]=='IC50'){d3$exp[i]<-1}
    if(d3$End[i]=='Dermal absorption'){
      d3$tox[i]<-1
      d3$t_vitro[i]<-1}
    if(d3$End[i] %in% c('Half-life (biotransformation)','Half life
(biotransformation)')){
      d3$tox[i]<-1
      if(d3$Spec[i] %in% c('Homo sapiens','Human')){d3$t_hmet[i]<-1}
      else{d3$t_amet[i]<-1}}
    if(d3$End[i] %in% c('Half-life (body)','Half-life (total body)')){
      d3$tox[i]<-1
      if(d3$Spec[i] %in% c('Homo sapiens','Human')){d3$t_habs[i]<-1}
      else{d3$t_aabs[i]<-1}}
    if(d3$End[i] %in% c('EC1.5','EC2','EC3')){
      if(d3$Type[i]=="in Vivo"){
        d3$animal[i]<-1
        d3$a_sens[i]<-1}
      else if(d3$Type[i]=="in Vitro"){
        d3$exp[i]<-1
        d3$e_conn[i]<-1}}
    if(d3$End[i]=="Reproductive toxicity"){
      d3$qual[i]<-1
      d3$ql_repro[i]<-1
      if(d3$TE[i]!="No relevant studies identified"){
        d3$animal[i]<-1
        d3$a_repro[i]<-1}}
    if(d3$Spec[i] %in% c('Peromyscus polionotus','Cattle','Sheep')){
      d3[i,2:ncolt]<-0
      d3$exp[i]<-1
    }
  }

```

```

}

d3$c<-paste0(as.character(d3$SMILES), as.character(d3$Ref))
d4<-d3[,c(4:ncolt, (ncolt+8))]

n1<- d4 %>% group_by(c, SMILES) %>%
  summarise(across(where(is.numeric), ~ max(.x)))

n2<-n1[,-1]
nsmi<- n2 %>% group_by(SMILES) %>%
  summarise_all(sum)
if(dim(d3)[1] != 0){
  all2<-
merge(nsmi, filter(tlist, is.na(tlist$CAS)), all.y=TRUE, by="SMILES") }
if(dim(d3)[1] == 0){all2<-filter(tlist, is.na(tlist$CAS))}

all<-bind_rows(all1, all2)
if(dim(all)[2] != (ncolt)){
commonNames <- names(n)[which(colnames(n) %in% colnames(tlist))]
commonNames <- commonNames[commonNames %notin% c("CAS", "SMILES")]
for(i in commonNames){
  left <- paste(i, ".x", sep="")
  right <- paste(i, ".y", sep="")
  all[is.na(all[left]), left] <- all[is.na(all[left]), right]
  all[right]<- NULL
  colnames(all)[colnames(all) == left] <- i
}

if(which(colnames(all)=="CID")==ncolt-2){all<-all[,c(ncolt-2, 1, (ncolt-
1):ncolt, 2:(ncolt-3))]}

all<-all[order(all$CID, all$CAS, na.last=FALSE),]
all[,5:ncolt][is.na(all[,5:ncolt])] <- 0

wb <- loadWorkbook(outfile)
writeData(wb, sheet = "OECD", all[1:ncolt], colNames = F, startRow=3)
saveWorkbook(wb, outfile, overwrite = T)

```