

Van de Wall et al.:

Comparing Translational Success Rates Across Medical Research Fields – A Combined Analysis of Literature and Clinical Trial Data

Supplementary Data

Appendix A – Search string

(Cancer Research OR Pharmacology OR Neuroscience) AND Human Studies AND Animal Studies AND Translation AND Review

Cancer (4,365,680 hits on 22-01-2021)

"Neoplasms"[mh:noexp] OR "Histiocytic Disorders, Malignant"[MeSH] OR "Leukemia"[MeSH] OR "Lymphangiosarcoma"[MeSH] OR "Lymphoma"[MeSH] OR "Neoplasms, Connective and Soft Tissue"[MeSH] OR "Neoplasms, Germ Cell and Embryonal"[MeSH] OR "Neoplasms, Glandular and Epithelial"[MeSH] OR "Neoplasms, Gonadal Tissue"[MeSH] OR "Neuroectodermal Tumors"[MeSH] OR "Multiple Myeloma"[MeSH] OR "Plasmacytoma"[MeSH] OR "Hemangiopericytoma"[MeSH] OR "Hemangiosarcoma"[MeSH] OR "Melanoma"[MeSH] OR "Neoplasms by Site"[MeSH] OR "Neoplasms, Experimental"[MeSH] OR "Neoplasms, Hormone-Dependent"[MeSH] OR "Neoplasms, Radiation Induced"[MeSH] OR "Neoplasms, Second Primary"[MeSH] OR "Neoplastic Processes"[MeSH] OR "Neoplastic Syndromes, Hereditary"[MeSH] OR "Choriocarcinoma"[MeSH] OR "Leukemia"[tiab] OR "Lymphangiosarcoma"[tiab] OR "Lymphoma"[tiab] OR "Neuroectodermal Tumors"[tiab] OR "Multiple Myeloma"[tiab] OR "Plasmacytoma"[tiab] OR "Hemangiopericytoma"[tiab] OR "Hemangiosarcoma"[tiab] OR "Melanoma"[tiab] OR "Neoplastic Processes"[tiab] OR "Choriocarcinoma"[tiab] OR "Neoplasia"[tiab] OR "Neoplasias"[tiab] OR "Neoplasm"[tiab] OR "Tumors"[tiab] OR "Tumor"[tiab] OR "Cancer"[tiab] OR "Cancers"[tiab] OR "Malignancy"[tiab] OR "Malignancies"[tiab] OR "Malignant Neoplasms"[tiab] OR "Malignant Neoplasm"[tiab] OR "Carcinoma"[tiab] OR "Carcinomas"[tiab] OR "Lump"[tiab] OR "Lumps"[tiab] OR "Tumour"[tiab] OR "Tumours"[tiab] OR "Sarcoma"[tiab] OR "Sarcomas"[tiab] OR "Tumefaction"[tiab] OR "Tumefactive"[tiab] OR "Malignant tumor"[tiab] OR "malignant tumors"[tiab] OR "Malignant Tumour"[tiab] OR "Malignant Tumours"[tiab] OR "Cancer Research"[tiab] OR "Tumor Research"[tiab] OR "Tumour Research"[tiab] OR "Melanoma"[tiab]

Pharmacology (6,729,613 hits on 22-01-2021)

"Pharmacology"[MeSH] OR Pharmacology[subheading] OR "Pharmacology"[tiab] OR "Pharmacologies"[tiab] OR "Pharmaceutics"[tiab] OR "Pharmacy"[tiab] OR "Pharmacological"[tiab] OR "Pharmaceuticals"[tiab] OR "Pharmacologically"[tiab] OR "Pharmaceutical"[tiab] OR "Pharma"[tiab] OR "Materia Medica"[tiab] OR "Pharmacologic"[tiab] OR "Pharmaceutics"[tiab] OR "Pharmaceutic"[tiab] OR "Pharmacotherapy"[tiab] OR "Pharm."[tiab] OR "Pharmaceut."[tiab] OR "Pharmacol."[tiab] OR "Pharmacological phenomena"[tiab] OR "Pharmacological phenomenon"[tiab] OR "Pharmacological processes"[tiab] OR "Pharmacological process"[tiab] OR "clinical pharmacology"[tiab] OR "Pharmaco Therapy"[tiab] OR "Pharmacodynamic"[tiab] OR "Pharmacokinetic"[tiab] OR "Toxicology"[tiab] OR "Adverse events"[tiab]

Neuroscience (530,280 hits on 22-01-2021)

"Neurosciences"[MeSH] OR "Neuroscience"[tiab] OR "Neurosciences"[tiab] OR "Neurology"[tiab] OR "Brain Science"[tiab] OR "Neurobiology"[tiab] OR "Neurophysiology"[tiab] OR "Cognitive Neuroscience"[tiab] OR "Neuroscientific"[tiab] OR "Brain Research"[tiab] OR "Neurologic"[tiab] OR "Neurological"[tiab] OR "NEURO"[tiab] OR "Neurosci"[tiab] OR "Neurodegenerative"[tiab] OR "Neurodegenerative disorder"[tiab] OR "Neurodegenerative disorders"[tiab] OR "Neurodevelopmental"[tiab] OR "Neurodevelopmental disorder"[tiab] OR "Neurodevelopmental disorders"[tiab]

Human Studies (19,744,856 hits on 22-01-2021)

"clinical study"[pt] OR "randomized controlled trial"[pt] OR "controlled clinical trial"[pt] OR humans[MeSH] OR "Clinical Trials as Topic"[MeSH] OR "intervention study"[tiab] OR "first-in-man"[tiab] OR "first in man"[tiab] OR trial[tiab] OR human[tiab] OR humans[tiab]

doi:10.14573/altex.2208261s

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.

Animal filter SYRCLE (1,710,962 hits on 22-01-2021)

"animal experimentation"[MeSH Terms] OR "models, animal"[MeSH Terms] OR "invertebrates"[MeSH Terms] OR "Animals"[Mesh:noexp] OR "animal population groups"[MeSH Terms] OR "chordata"[MeSH Terms:noexp] OR "chordata, nonvertebrate"[MeSH Terms] OR "vertebrates"[MeSH Terms:noexp] OR "amphibians"[MeSH Terms] OR "birds"[MeSH Terms] OR "fishes"[MeSH Terms] OR "reptiles"[MeSH Terms] OR "mammals"[MeSH Terms:noexp] OR "primates"[MeSH Terms:noexp] OR "artiodactyla"[MeSH Terms] OR "carnivora"[MeSH Terms] OR "cetacea"[MeSH Terms] OR "chiroptera"[MeSH Terms] OR "elephants"[MeSH Terms] OR "hyraxes"[MeSH Terms] OR "insectivora"[MeSH Terms] OR "lagomorpha"[MeSH Terms] OR "marsupialia"[MeSH Terms] OR "monotremata"[MeSH Terms] OR "perissodactyla"[MeSH Terms] OR "rodentia"[MeSH Terms] OR "scandentia"[MeSH Terms] OR "sirenia"[MeSH Terms] OR "xenarthra"[MeSH Terms] OR "haplorhini"[MeSH Terms:noexp] OR "strepsirhini"[MeSH Terms] OR "platyrrhini"[MeSH Terms] OR "tarsii"[MeSH Terms] OR "catarrhini"[MeSH Terms:noexp] OR "cercopithecidae"[MeSH Terms] OR "hylobatidae"[MeSH Terms] OR "hominidae"[MeSH Terms:noexp] OR "gorilla gorilla"[MeSH Terms] OR "pan paniscus"[MeSH Terms] OR "pan troglodytes"[MeSH Terms] OR "pongo pygmaeus"[MeSH Terms] OR ((animals[tiab] OR animal[tiab] OR mice[tiab] OR mus[tiab] OR mouse[tiab] OR murine[tiab] OR woodmouse[tiab] OR rats[tiab] OR rat[tiab] OR murinae[tiab] OR muridae[tiab] OR cottonrat[tiab] OR cottonrats[tiab] OR hamster[tiab] OR hamsters[tiab] OR cricetinae[tiab] OR rodentia[tiab] OR rodent[tiab] OR rodents[tiab] OR pigs[tiab] OR pig[tiab] OR swine[tiab] OR swines[tiab] OR piglets[tiab] OR piglet[tiab] OR boar[tiab] OR boars[tiab] OR "sus scrofa"[tiab] OR ferrets[tiab] OR ferret[tiab] OR polecat[tiab] OR polecats[tiab] OR "mustela putorius"[tiab] OR "guinea pigs"[tiab] OR "guinea pig"[tiab] OR cavia[tiab] OR callithrix[tiab] OR marmoset[tiab] OR marmosets[tiab] OR cebuella[tiab] OR hapale[tiab] OR octodon[tiab] OR chinchilla[tiab] OR chinchillas[tiab] OR gerbillinae[tiab] OR gerbil[tiab] OR gerbils[tiab] OR jird[tiab] OR jirds[tiab] OR merione[tiab] OR meriones[tiab] OR rabbits[tiab] OR rabbit[tiab] OR hares[tiab] OR hare[tiab] OR diptera[tiab] OR flies[tiab] OR fly[tiab] 21 OR dipteral[tiab] OR drosophila[tiab] OR drosophilidae[tiab] OR cats[tiab] OR cat[tiab] OR carus[tiab] OR felis[tiab] OR nematoda[tiab] OR nematode[tiab] OR nematodes[tiab] OR sipunculida[tiab] OR dogs[tiab] OR dog[tiab] OR canine[tiab] OR canines[tiab] OR canis[tiab] OR sheep[tiab] OR sheeps[tiab] OR mouflon[tiab] OR mouflons[tiab] OR ovis[tiab] OR goats[tiab] OR goat[tiab] OR capra[tiab] OR capras[tiab] OR rupicapra[tiab] OR rupicapras[tiab] OR chamois[tiab] OR haplorhini[tiab] OR monkey[tiab] OR monkeys[tiab] OR anthropoidea[tiab] OR anthropoids[tiab] OR saguinus[tiab] OR tamarin[tiab] OR tamarins[tiab] OR leontopithecus[tiab] OR hominidae[tiab] OR ape[tiab] OR apes[tiab] OR "pan paniscus"[tiab] OR bonobo[tiab] OR bonobos[tiab] OR "pan troglodytes"[tiab] OR gibbon[tiab] OR gibbons[tiab] OR siamang[tiab] OR siamangs[tiab] OR nomascus[tiab] OR symphalangus[tiab] OR chimpanzee[tiab] OR chimpanzees[tiab] OR prosimian[tiab] OR prosimians[tiab] OR "bush baby"[tiab] OR bush babies[tiab] OR galagos[tiab] OR galago[tiab] OR pongidae[tiab] OR gorilla[tiab] OR gorillas[tiab] OR "pongo pygmaeus"[tiab] OR orangutan[tiab] OR orangutans[tiab] OR lemur[tiab] OR lemurs[tiab] OR lemuriidae[tiab] OR horse[tiab] OR horses[tiab] OR equus[tiab] OR cow[tiab] OR calf[tiab] OR bull[tiab] OR chicken[tiab] OR chickens[tiab] OR gallus[tiab] OR quail[tiab] OR bird[tiab] OR birds[tiab] OR quails[tiab] OR poultry[tiab] OR poultries[tiab] OR fowl[tiab] OR fowls[tiab] OR reptile[tiab] OR reptilia[tiab] OR reptiles[tiab] OR snakes[tiab] OR snake[tiab] OR lizard[tiab] OR lizards[tiab] OR alligator[tiab] OR alligators[tiab] OR crocodile[tiab] OR crocodiles[tiab] OR turtle[tiab] OR turtles[tiab] OR amphibian[tiab] OR amphibians[tiab] OR amphibia[tiab] OR frog[tiab] OR frogs[tiab] OR bombina[tiab] OR salientia[tiab] OR toad[tiab] OR toads[tiab] OR "epidalea calamita"[tiab] OR salamander[tiab] OR salamanders[tiab] OR eel[tiab] OR eels[tiab] OR fish[tiab] OR fishes[tiab] OR pisces[tiab] OR catfish[tiab] OR catfishes[tiab] OR siluriformes[tiab] OR arius[tiab] OR heteropneustes[tiab] OR sheatfish[tiab] OR perch[tiab] OR perches[tiab] OR percidae[tiab] OR perca[tiab] OR trout[tiab] OR trouts[tiab] OR char[tiab] OR chars[tiab] OR salvelinus[tiab] OR minnow[tiab] OR cyprinidae[tiab] OR carps[tiab] OR carp[tiab] OR zebrafish[tiab] OR zebrafishes[tiab] OR goldfish[tiab] OR goldfishes[tiab] OR guppy[tiab] OR guppies[tiab] OR chub[tiab] OR chubs[tiab] OR tinca[tiab] OR barbels[tiab] OR barbuis[tiab] OR pimphales[tiab] OR promelas[tiab] OR "poecilia reticulata"[tiab] OR mullet[tiab] OR mullets[tiab] OR eel[tiab] OR eels[tiab] OR seahorse[tiab] OR seahorses[tiab] OR mugil curema[tiab] OR atlantic cod[tiab] OR shark[tiab] OR sharks[tiab] OR catshark[tiab] OR anguilla[tiab] OR salmonid[tiab] OR salmonids[tiab] OR whitefish[tiab] OR whitefishes[tiab] OR salmon[tiab] OR salmon[tiab] OR sole[tiab] OR solea[tiab] OR lamprey[tiab] OR lampreys[tiab] OR pumpkinseed[tiab] OR sunfish[tiab] OR sunfishes[tiab] OR tilapia[tiab] OR tilapias[tiab] OR turbot[tiab] OR turbots[tiab] OR flatfish[tiab] OR flatfishes[tiab] OR sciuridae[tiab] OR squirrel[tiab] OR squirrels[tiab] OR chipmunk[tiab] OR chipmunks[tiab] OR suslik[tiab] OR susliks[tiab] OR vole[tiab] OR voles[tiab] OR lemming[tiab] OR lemmings[tiab] OR muskrat[tiab] OR muskrats[tiab] OR lemmus[tiab] OR otter[tiab] OR otters[tiab] OR marten[tiab] OR martens[tiab] OR martes[tiab] OR weasel[tiab] OR badger[tiab] OR badgers[tiab] OR ermine[tiab] OR mink[tiab] OR minks[tiab] OR sable[tiab] OR sables[tiab] OR gulo[tiab] OR gulos[tiab] OR wolverine[tiab] OR wolverines[tiab] OR mustela[tiab] OR llama[tiab] OR llamas[tiab] OR alpaca[tiab] OR alpacas[tiab] OR camelid[tiab] OR camelids[tiab] OR guanaco[tiab] OR guanacos[tiab] OR chiroptera[tiab] OR chiropteras[tiab] OR bat[tiab] OR bats[tiab] OR fox[tiab] OR foxes[tiab] OR iguana[tiab] OR iguanas[tiab] OR xenopus laevis[tiab] OR parakeet[tiab] OR parakeets[tiab] OR parrot[tiab] OR parrots[tiab] OR donkey[tiab] OR donkeys[tiab] OR mule[tiab] OR mules[tiab] OR zebra[tiab] OR zebras[tiab] OR shrew[tiab] OR shrews[tiab] OR bison[tiab] OR bisons[tiab] OR buffalo[tiab] OR buffaloes[tiab] OR deer[tiab] OR deers[tiab] OR bear[tiab] OR bears[tiab] OR panda[tiab] OR pandas[tiab] OR "wild hog"[tiab] OR "wild boar"[tiab] OR fitchew[tiab] OR fitch[tiab] OR beaver[tiab] OR beavers[tiab] OR

jerboa[Tiab] OR jerboas[Tiab] OR capybara[Tiab] OR capybaras[Tiab] OR canine [tiab] OR bovine [tiab] OR porcine [tiab] OR hog [tiab] OR hogs [tiab]) NOT medline[sb]

Translation (7,944,723 hits on 22-01-2021)

"Translational Medical Research"[MESH] OR "Translation"[tiab] OR Translations[tiab] OR Translational[tiab] OR "Bench to Bedside"[tiab] OR Extrapol*[tiab] OR validat*[tiab] OR compar*[tiab] OR correlat*[tiab] OR predicta*[tiab] OR predicti*[tiab] OR predictor*[tiab]

Review (6,229,580 hits on 22-01-2021)

review[pt] OR letter[pt] OR editorial[pt] OR "systematic review"[pt] OR "Comparative Study"[pt]

Appendix B – R script

B.1 Custom function, Utils.R

```
1 #Functions
2
3 #Calculate percentages
4 Percentage_calc <- function(DataFrame) {
5 Result <- ((DataFrame$N[[2]]/(DataFrame$N[[2]]+DataFrame$N[[1]]))*100
6 return(Result)
7 }
```

B.2 Overall results

```
1 #This script investigates the success and trial design across all included clinical
  trials from the WHO clinical trial registry
2
3 #Clear workspace
4 rm(list = ls(all.names = TRUE))
5
6 #Packages
7 library(readxl)
8 library(dplyr)
9 library(sjmisc)
10 source("Utils.R")
11
12 #Import Data Set
13 Data<-read_xlsx('Data/Preliminary_Data.xlsx')
14
15 #ICD Data Frame
16 .ICD_Codes <- data.frame(doc_id = Data$Id, code = Data$`ICD-10 Code`, success = Data$
  Trial_Success`, Randomised = Data$Randomized)
17
18 #Overall Results Percentage
19 .Termination <- dplyr::count(Data, Terminated)
20 .tot_perc_term <- Percentage_calc(.Termination)
21 Overall_Results <- data.frame(.tot_perc_term)
22 .succes_overall <- dplyr::count(.ICD_Codes, success)
23 Overall_Results$Tot_perc_suc <- Percentage_calc(.succes_overall)
24 .randomised_overall <- dplyr::count(Data, Randomized)
25 Overall_Results$Tot_perc_rand <- Percentage_calc(.randomised_overall)
26 .single_overall <- dplyr::count(Data, Single_Blind)
27 Overall_Results$Tot_perc_single <- Percentage_calc(.single_overall)
28 .double_overall <- dplyr::count(Data, Double_Blind)
29 Overall_Results$Tot_perc_double <- Percentage_calc(.double_overall)
30 .triple_overall <- dplyr::count(Data, Triple_blind)
31 Overall_Results$Tot_perc_triple <- Percentage_calc(.triple_overall)
32 Overall_Results$Tot_perc_blind <- Overall_Results$Tot_perc_triple + Overall_Results$
  Tot_perc_double + Overall_Results$Tot_perc_single
33 .control_overall <- dplyr::count(Data, Controlled)
34 Overall_Results$Tot_perc_control <- Percentage_calc(.control_overall)
35 .RCT <- subset(Data, Controlled == TRUE & Randomized == TRUE)
36 Overall_Results$tot_perc_RCT <- (nrow(.RCT)/nrow(Data))*100
```

```

37
38 #Termination reasons
39 Termination_Reason <- dplyr::count(Data, Notes)
40
41 #Efficacy Trials
42 Efficacy <- subset(Data, Safety == FALSE)
43
44 #Randomised vs non-randomised, Efficacy
45 .Randomised <- subset(Efficacy, Randomized == TRUE)
46 .Non_Randomised <- subset (Efficacy, Randomized == FALSE)
47 .success_randomised <- dplyr::count(.Randomised, Trial_Success)
48 .success_non_randomised <- dplyr::count(.Non_Randomised, Trial_Success)
49 .Randomised_Success <- Percentage_calc(.success_randomised)
50 .Non_Randomised_Success <- Percentage_calc(.success_non_randomised)
51
52 #Controlled vs Non-Controlled, Efficacy
53 .Controlled <- subset(Efficacy, Controlled == TRUE)
54 .Non_Controlled <- subset (Efficacy, Controlled == FALSE)
55 .success_Controlled <- dplyr::count(.Controlled, Trial_Success)
56 .success_non_Controlled <- dplyr::count(.Non_Controlled, Trial_Success)
57 .Controlled_Success <- Percentage_calc(.success_Controlled)
58 .Non_Controlled_Success <- Percentage_calc(.success_non_Controlled)
59
60 #RCT vs non-RCT, Efficacy
61 .non_RCT <- subset(Efficacy, Controlled == FALSE & Randomized == FALSE)
62 .RCT_efficacy <- subset(Efficacy, Controlled == TRUE & Randomized == TRUE)
63 .success_non_RCT <- dplyr::count(.non_RCT, .non_RCT$Trial_Success)
64 .Non_RCT_Success <- Percentage_calc(.success_non_RCT)
65 .success_RCT <- dplyr::count(.RCT_efficacy, Trial_Success)
66 .RCT_Success <- Percentage_calc(.success_RCT)
67
68 #Add Efficacy Results to a dataframe
69 Efficacy_Overall_Results <- data.frame('Controlled' = .Controlled_Success, 'Not
Controlled' = .Non_Controlled_Success,
70 'Randomised' = .Randomised_Success, 'Not
Randomised' = .Non_Randomised_Success,
71 'RCT' = .RCT_Success, 'Not RCT' = .
Non_RCT_Success)

```

B.3 Grouping

```

1 #This script automatically groups all clinical trials based on their assigned ICD-10
code with a minimum group size of the chosen threshold
2 #Here, threshold = 50
3
4 #Clear workspace
5 rm(list = ls(all.names = TRUE))
6
7 #Packages
8 library(readxl)
9 library(writexl)
10 library(dplyr)
11 library(sjmisc)
12
13 #Import Data Set
14 Data<-read_xlsx('Data/Preliminary_Data.xlsx')
15
16 #ICD Data Frame
17 .ICD_Codes <- data.frame(doc_id = Data$Id, code = Data`ICD-10 Code`, success = Data$
Trial_Success`, Randomised = Data$Randomized)
18
19 #Count Codes
20 Code_Counts <- dplyr::count(.ICD_Codes, code, sort=FALSE)
21
22 #Get codes with sufficient trials

```

```

23 Sufficient_Base <- subset(Code_Counts, n >= 50)
24
25 #Group insufficient cases on first three characters
26 .old <- Code_Counts$code[1]
27 .nums <- 0
28 .total <- 0
29 group_names <- rep(NA, nrow(Code_Counts))
30 .treshhold <- 50
31
32 Grouped_Codes <- data.frame(matrix(ncol = 2, nrow = 0))
33 .x <- c("group", "total")
34 colnames(Grouped_Codes) <- .x
35
36 #go trough each element
37 for(i in 2:(nrow(Code_Counts))){
38 if(Code_Counts$n[i] < .treshhold && Code_Counts$n[i-1] < .treshhold){
39 #check if code is similar as the one above it
40 if( str_contains(Code_Counts$code[i], substr(.old, 0, 3)) ){
41 # if it is the first element that is similar also take the one above
42 if (.nums == 0){
43 .total = .total + Code_Counts$n[i-1]
44 group_names[i-1] = substr(.old, 0, 3)
45 }
46
47 group_names[i] = substr(.old, 0, 3)
48 .total = .total + Code_Counts$n[i]
49 .nums = .nums + 1
50 }
51 else{
52 if (.nums != 0){
53 y <- data.frame(group=substr(.old, 0, 3), .total = .total)
54 Grouped_Codes <- rbind(Grouped_Codes, y)
55 }
56 #Reset Variables
57 .nums = 0
58 .total = 0
59
60 #Add standard name to the group
61 group_names[i] = Code_Counts$code[i]
62 #Current code in old variable for use in the next loop
63 .old = Code_Counts$code[i]
64
65 }
66 }
67 if(is.na(group_names[i])){
68 group_names[i] = Code_Counts$code[i]
69 }
70 }
71 #If, after original grouping, the group size is still smaller than treshhold: replace
group with just the first letter
72 for(i in 1:(nrow(Grouped_Codes))){
73 if(Grouped_Codes$total[i]<.treshhold){
74 Grouped_Codes$group[i] = substr(Grouped_Codes$group[i], 0, 1)
75 }
76 }
77 #group Grouped_Codes to sum together all entries by group
78 Grouped_Codes = aggregate(Grouped_Codes$total, by=list(Category = Grouped_Codes$group
), FUN=sum)
79 colnames(Grouped_Codes) <- .x
80
81 #group everything that's still too small
82 for(i in 1:(nrow(Grouped_Codes))){
83 if(Grouped_Codes$total[i]<.treshhold){
84 Grouped_Codes$group[i] = "Others"

```

```

85 }
86 }
87 Grouped_Codes = aggregate(Grouped_Codes$total, by=list(Category = Grouped_Codes$group)
, FUN=sum)
88 colnames(Grouped_Codes) <- .x
89
90 #put groupname
91 Code_Counts$groups=group_names
92
93 #combine Sufficient_base with Grouped_codes
94 colnames(Sufficient_Base)<-colnames(Grouped_Codes)
95 All_Groups <- rbind(Grouped_Codes, Sufficient_Base)
96
97 #Export
98 All_Groups_Export <- write_xlsx(All_Groups, "data//All_Groups.xlsx")

```

B.4 Subset analysis

```

1 #This script extracts the several groups (created in the script "Grouping.R") from
the full dataset
2 #and calculates the success within these groups
3
4 #Step 1: Run Grouping.R to retrieve the subgroups
5 #Step 2: Add disease name column to "All_Groups.xlsx" this file is an output from
"Grouping.R"
6
7 #Clear workspace
8 rm(list = ls(all.names = TRUE))
9
10 #Packages
11 library(readxl)
12 library(dplyr)
13 library(sjmisc)
14 library(stringr)
15 source("Utils.R")
16
17 #Import Data Set
18 Data<-read_xlsx('Data/Preliminary_Data.xlsx')
19 Subgroups <- read_xlsx('Data/All_Groups.xlsx')
20
21 #Create Empty Dataframes
22 Subgroup_Success <- data.frame(matrix(ncol = 2, nrow = 0))
23 Subgroup_Termination <- data.frame(matrix(ncol = 2, nrow = 0))
24 #Create malleable dataframes
25 .Data_Subset <- Data
26 .Subgroups_Subset <- Subgroups
27
28 for(i in 1:(nrow(Subgroups))){
29 #If code has 5 characters
30 if (nchar(Subgroups$group[[i]]) == 5){
31 #subset code
32 .a <- subset(.Data_Subset, .Data_Subset$`ICD-10 Code` == Subgroups$group[[i]])
33 #Count Success and Termination
34 .b <- dplyr::count(.a, Trial_Success)
35 .c <- dplyr::count(.a, Terminated)
36 #Create dataframes with percentage positive outcomes and terminations
37 .z <- data.frame(Field = Subgroups$disease[[i]], Success = Percentage_calc(.b))
38 .y <- data.frame(Field = Subgroups$disease[[i]], Termination = Percentage_calc(.c
))
39 #Add percentages to full overview dataframe
40 Subgroup_Success <- rbind(Subgroup_Success, .z)
41 Subgroup_Termination <- rbind(Subgroup_Termination, .y)
42 #Delete Used data from primary dataframes
43 .Data_Subset <- subset(.Data_Subset, .Data_Subset$`ICD-10 Code` != Subgroups$group
[[i]])

```

```

44 .Subgroups_Subset <- subset(.Subgroups_Subset, .Subgroups_Subset$group !=
Subgroups$group[[i]])
45 }
46 }
47
48 #Create malleable dataframes
49 .Subgroups_Subset2 <- .Subgroups_Subset
50
51 for(i in 1:(nrow(.Subgroups_Subset))){
52 #If code has 3 characters
53 if (nchar(.Subgroups_Subset$group[[i]]) == 3){
54 #subset code
55 .a <- subset(.Data_Subset, grepl(paste('^', .Subgroups_Subset$group[[i]], sep = ""
), .Data_Subset$`ICD-10 Code`))
56 #Count Success
57 .b <- dplyr::count(.a, Trial_Success)
58 .c <- dplyr::count(.a, Terminated)
59 #Create dataframes with percentage positive outcomes and terminations
60 .z <- data.frame(Field = .Subgroups_Subset$disease[[i]], Success = Percentage_calc
(.b))
61 .y <- data.frame(Field = .Subgroups_Subset$disease[[i]], Termination =
Percentage_calc(.c))
62 #Add percentages to full overview dataframe
63 Subgroup_Success <- rbind(Subgroup_Success, .z)
64 Subgroup_Termination <- rbind(Subgroup_Termination, .y)
65 #Delete Used data from primary dataframes
66 .Data_Subset <- subset(.Data_Subset, !str_detect(.Data_Subset$`ICD-10 Code`, .
Subgroups_Subset$group[[i]])
67 .Subgroups_Subset2 <- subset(.Subgroups_Subset2, .Subgroups_Subset2$group != .
Subgroups_Subset$group[[i]])
68 }
69 }
70
71 #Create Empty Dataframes
72 Chapter_Success <- data.frame(matrix(ncol = 2, nrow = 0))
73 Chapter_Termination <- data.frame(matrix(ncol = 2, nrow = 0))
74 #Create malleable dataframes
75 .Subgroups_Subset3 <- .Subgroups_Subset2
76
77 for(i in 1:(nrow(.Subgroups))){
78 #If code has 1 character
79 if (nchar(.Subgroups$group[[i]]) == 1){
80 #subset code
81 .a <- subset(Data, grepl(paste('^', .Subgroups$group[[i]], sep = ""), Data
$`ICD-10 Code`))
82 #Count Success and Termination
83 .b <- dplyr::count(.a, Trial_Success)
84 .c <- dplyr::count(.a, Terminated)
85 #Create dataframes with percentage positive outcomes and terminations
86 .z <- data.frame(Field = Subgroups$disease[[i]], Success =
Percentage_calc(.b))
87 .y <- data.frame(Field = Subgroups$disease[[i]], Termination =
Percentage_calc(.c))
88 #Add percentages to full overview dataframe
89 Chapter_Success <- rbind(Chapter_Success, .z)
90 Chapter_Termination <- rbind(Chapter_Termination, .y)
91 #Delete Used data from primary dataframes
92 .Data_Subset <- subset(.Data_Subset, !str_detect(.Data_Subset$`ICD-10 Code`, .
Subgroups_Subset2$group[[i]])
93 }
94 }
95
96 #Calculate percentages of remaining non-grouped trials
97 #Count Success and Termination

```

```
98 .b <- dplyr::count(.Data_Subset, Trial_Success)
99 .c <- dplyr::count(.Data_Subset, Terminated)
100 #Create dataframes with percentage positive outcomes and terminations
101 .z <- data.frame(Field = 'Other', Success = Percentage_calc(.b))
102 .y <- data.frame(Field = 'Other', Termination = Percentage_calc(.c))
103 #Add percentages to full overview dataframe
104 Chapter_Success <- rbind(Chapter_Success, .z)
105 Chapter_Termination <- rbind(Chapter_Termination, .y)
```