TSAR: Clinical EDA 2

Joshua Burkhart

August 18, 2016

Scan for Missing Data...

Fix Missing Data...

Find max/median log symptom score by time hours with study & virus summarizations

```
train_clinicl_df %>%
  dplyr::group_by(STUDYID) %>%
  dplyr::mutate(JACKSON_SCORE = 2^LOGSYMPTSCORE_SC3) %>%
  dplyr::summarise(jackson_mean=mean(JACKSON_SCORE),
                   jackson_med=median(JACKSON_SCORE),
                   log_mean=mean(LOGSYMPTSCORE_SC3),
                   log_med=median(LOGSYMPTSCORE_SC3)) %>%
  pander(split.cells = 50, split.table = Inf)
```

STUDYID	jackson_mean	jackson_med	log_mean	log_med
DEE1 RSV	1.946142	1.517882	0.8648848	0.6020600
DEE2 H3N2	1.888931	1.870067	0.8009716	0.9030900
DEE3 H1N1	1.910201	1.997888	0.8290704	0.9978176
DEE4X H1N1	1.728765	1.623345	0.6900376	0.6989700
DEE5 H3N2	1.954856	1.870067	0.9109137	0.9030900
Rhinovirus Duke	2.127819	2.058214	1.0220678	1.0413927
Rhinovirus UVA	1.863816	1.937562	0.8307942	0.9542425

```
train_clinicl_df %>%
  dplyr::group_by(STUDYID,TIMEHOURS) %>%
  dplyr::mutate(JACKSON_SCORE = 2^LOGSYMPTSCORE_SC3) %>%
  dplyr::summarise(jackson_mean=mean(JACKSON_SCORE),
                   jackson_med=median(JACKSON_SCORE),
                   log_mean=mean(LOGSYMPTSCORE_SC3),
                   log_med=median(LOGSYMPTSCORE_SC3)) %>%
  pander(split.cells = 50, split.table = Inf)
```

STUDYID	TIMEHOURS	jackson_mean	jackson_med	log_mean	log_med
DEE1 RSV	-24.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	0.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	5.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	12.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	21.5	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	29.0	1.943619	1.517882	0.8579652	0.6020600
DEE1 RSV	36.0	1.952080	1.693975	0.8690260	0.7525750

STUDYID	TIMEHOURS	jackson_mean	jackson_med	log_mean	log_med
DEE1 RSV	45.5	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	53.0	1.909786	1.517882	0.8377959	0.6020600
DEE1 RSV	60.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	69.5	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	77.0	1.872511	1.517882	0.8204278	0.6020600
DEE1 RSV	84.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	93.5	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	101.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	108.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	117.5	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	125.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	132.0	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	141.5	1.952080	1.693975	0.8690260	0.7525750
DEE1 RSV	165.5	1.952080	1.693975	0.8690260	0.7525750
DEE2 H3N2	-24.0	1.912537	1.870067	0.8138094	0.9030900
DEE2 H3N2	0.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	5.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	12.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	21.5	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	29.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	36.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	45.5	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	53.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	60.0	1.823068	1.669139	0.7348510	0.7385606
DEE2 H3N2	69.5	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	77.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	84.0	1.847337	1.792499	0.7690592	0.8406206
DEE2 H3N2	93.5	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	101.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	108.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	117.5	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	125.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	132.0	1.895526	1.870067	0.8070541	0.9030900
DEE2 H3N2	141.5	1.845432	1.792499	0.7680348	0.8406206
DEE2 H3N2	165.5	1.895526	1.870067	0.8070541	0.9030900
DEE3 H1N1	-24.0	1.893300	1.997888	0.8103086	0.9978176
DEE3 H1N1	0.0	1.950784	2.058214	0.8663983	1.0413927
DEE3 H1N1	5.0	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	12.0	1.888772	1.937562	0.8108003	0.9542425
DEE3 H1N1	21.5	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	29.0	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	36.0	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	45.5	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	53.0	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	60.0	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	69.5	1.910021	2.058214	0.8249094	1.0413927
DEE3 H1N1	77.0	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	84.0	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	93.5	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	101.0	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	108.0	1.911168	1.997888	0.8302983	0.9978176

STUDYID	TIMEHOURS	jackson_mean	jackson_med	log_mean	log_med
DEE3 H1N1	117.5	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	125.0	1.911168	1.997888	0.8302983	0.9978176
DEE3 H1N1	141.5	1.892245	1.937562	0.8129005	0.9542425
DEE3 H1N1	165.5	1.911168	1.997888	0.8302983	0.9978176
DEE4X H1N1	-24.0	1.797457	1.623345	0.7415125	0.6989700
DEE4X H1N1	0.0	1.635479	1.623345	0.6336535	0.6989700
DEE4X H1N1	5.0	1.750248	1.623345	0.7134378	0.6989700
DEE4X H1N1	12.0	1.782820	1.623345	0.7349211	0.6989700
DEE4X H1N1	21.5	1.840985	1.664763	0.7521482	0.7156819
DEE4X H1N1	29.0	1.718502	1.623345	0.7158782	0.6989700
DEE4X H1N1	36.0	1.750248	1.623345	0.7134378	0.6989700
DEE4X H1N1	45.5	1.712601	1.623345	0.6770282	0.6989700
DEE4X H1N1	53.0	1.704223	1.507654	0.6602212	0.5880456
DEE4X H1N1	60.0	1.690105	1.507654	0.6493067	0.5880456
DEE4X H1N1	69.5	1.758325	1.507654	0.7017185	0.5880456
DEE4X H1N1	77.0	1.782820	1.623345	0.7349211	0.6989700
DEE4X H1N1	84.0	1.750248	1.623345	0.7134378	0.6989700
DEE4X H1N1	93.5	1.750248	1.623345	0.7134378	0.6989700
DEE4X H1N1	101.0	1.750248	1.623345	0.7134378	0.6989700
DEE4X H1N1	108.0	1.721527	1.507654	0.6748340	0.5880456
DEE4X H1N1	117.5	1.712601	1.623345	0.6770282	0.6989700
DEE4X H1N1	125.0	1.712601	1.623345	0.6770282	0.6989700
DEE4X H1N1	132.0	1.604651	1.507654	0.5758075	0.5880456
DEE4X H1N1	141.5	1.599686	1.507654	0.5887647	0.5880456
DEE4X H1N1	165.5	1.704223	1.507654	0.6602212	0.5880456
DEE5 H3N2	-30.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	2.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	10.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	18.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	26.0	1.941063	1.870067	0.8967332	0.9030900
DEE5 H3N2	34.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	42.0	1.981854	1.903815	0.9284174	0.9286662
DEE5 H3N2	50.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	58.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	66.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	74.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	82.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	90.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	98.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	106.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	114.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	122.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	130.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	138.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	146.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	162.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	170.0	1.954276	1.870067	0.9107676	0.9030900
DEE5 H3N2	672.0	1.954276	1.870067	0.9107676	0.9030900
Rhinovirus Duke	-24.0	2.144866	2.202307	1.0355526	1.1359208
Rhinovirus Duke	0.0	2.097366	2.058214	1.0031317	1.0413927
Rhinovirus Duke	4.0	2.144866	2.202307	1.0355526	1.1359208

	TIMEHOURS				
STUDYID		jackson_mean	jackson_med	log_mean	log_med
Rhinovirus Duke	8.0	2.088657	2.058214	0.9983856	1.0413927
Rhinovirus Duke	12.0	2.147889	2.202307	1.0408019	1.1359208
Rhinovirus Duke	16.0	2.122238	2.058214	1.0220052	1.0413927
Rhinovirus Duke	20.0	2.122238	2.058214	1.0220052	1.0413927
Rhinovirus Duke	24.0	2.097293	2.058214	1.0058534	1.0413927
Rhinovirus Duke	30.0	2.144866	2.202307	1.0355526	1.1359208
Rhinovirus Duke	36.0	2.122238	2.058214	1.0220052	1.0413927
Rhinovirus Duke	42.0	2.122238	2.058214	1.0220052	1.0413927
Rhinovirus Duke	48.0	2.122238	2.058214	1.0220052	1.0413927
Rhinovirus Duke	60.0	2.097366	2.058214	1.0031317	1.0413927
Rhinovirus Duke	72.0	2.125250	2.058214	1.0210235	1.0413927
Rhinovirus Duke	84.0	2.157074	2.346400	1.0358571	1.2304489
Rhinovirus Duke	96.0	2.081956	2.202307	0.9782097	1.1359208
Rhinovirus Duke	108.0	2.192423	2.535768	1.0501051	1.3424227
Rhinovirus Duke	120.0	2.164135	2.441084	1.0354617	1.2864358
Rhinovirus Duke	136.0	2.158317	2.441084	1.0306798	1.2864358
Rhinovirus UVA	-24.0	1.892255	1.968781	0.8586616	0.9771213
Rhinovirus UVA	0.0	1.863689	1.968781	0.8301018	0.9771213
Rhinovirus UVA	4.0	1.863689	1.968781	0.8301018	0.9771213
Rhinovirus UVA	8.0	1.906745	2.000000	0.8700497	1.0000000
Rhinovirus UVA	12.0	1.863689	1.968781	0.8301018	0.9771213
Rhinovirus UVA	16.0	1.863689	1.968781	0.8301018	0.9771213
Rhinovirus UVA	20.0	1.863689	1.968781	0.8301018	0.9771213
Rhinovirus UVA	24.0	1.863689	1.968781	0.8301018	0.9771213
Rhinovirus UVA	30.0	1.863689	1.968781	0.8301018	0.9771213
Rhinovirus UVA	36.0	1.816881	1.903815	0.7911283	0.9286662
Rhinovirus UVA	42.0	1.850576	1.937562	0.8169924	0.9542425
Rhinovirus UVA	48.0	1.826519	1.937562	0.8021215	0.9542425
Rhinovirus UVA	72.0	1.863689	1.968781	0.8301018	0.9771213
Rhinovirus UVA	96.0	1.896935	2.000000	0.8579477	1.0000000
Rhinovirus UVA	120.0	1.863689	1.968781	0.8301018	0.9771213

```
train_clinicl_df %>%
  add_virus_col() %>%
  dplyr::group_by(VIRUS) %>%
  dplyr::mutate(JACKSON_SCORE = 2^LOGSYMPTSCORE_SC3) %>%
  dplyr::summarise(jackson_mean=mean(JACKSON_SCORE),
                   jackson_med=median(JACKSON_SCORE),
                   log_mean=mean(LOGSYMPTSCORE_SC3),
                   log_med=median(LOGSYMPTSCORE_SC3)) %>%
  pander(split.cells = 50, split.table = Inf)
```

VIRUS	jackson_mean	jackson_med	log_mean	log_med
H1N1	1.853584	1.796387	0.7856857	0.845098
H3N2	1.919147	1.870067	0.8513617	0.903090
Rhinovirus	2.004162	2.058214	0.9324761	1.041393
RSV	1.946142	1.517882	0.8648848	0.602060

```r
train_clinicl_df %>%
  add_virus_col() %>%
  dplyr::group_by(VIRUS,TIMEHOURS) %>%
  dplyr::mutate(JACKSON_SCORE = 2^LOGSYMPTSCORE_SC3) %>%
  dplyr::summarise(jackson_mean=mean(JACKSON_SCORE),
                   jackson_med=median(JACKSON_SCORE),
                   log_mean=mean(LOGSYMPTSCORE_SC3),
                   log_med=median(LOGSYMPTSCORE_SC3)) %>%
  pander(split.cells = 50, split.table = Inf)
```

VIRUS	TIMEHOURS	jackson_mean	jackson_med	log_mean	log_med
H1N1	-24.0	1.863349	1.866974	0.7888098	0.8996703
H1N1	0.0	1.848774	1.796387	0.7910985	0.8450980
H1N1	5.0	1.857528	1.796387	0.7913448	0.8450980
H1N1	12.0	1.854493	1.796387	0.7862511	0.8450980
H1N1	21.5	1.893623	1.937562	0.8107608	0.9542425
H1N1	29.0	1.867663	1.796387	0.8044615	0.8450980
H1N1	36.0	1.857528	1.796387	0.7913448	0.8450980
H1N1	45.5	1.848762	1.796387	0.7821277	0.8450980
H1N1	53.0	1.850302	1.755659	0.7802756	0.8116246
H1N1	60.0	1.846150	1.755659	0.7770655	0.8116246
H1N1	69.5	1.864052	1.796387	0.7875789	0.8450980
H1N1	77.0	1.870830	1.796387	0.8003226	0.8450980
H1N1	84.0	1.857528	1.796387	0.7913448	0.8450980
H1N1	93.5	1.857528	1.796387	0.7913448	0.8450980
H1N1	101.0	1.857528	1.796387	0.7913448	0.8450980
H1N1	108.0	1.855392	1.796387	0.7845735	0.8450980
H1N1	117.5	1.848762	1.796387	0.7821277	0.8450980
H1N1	125.0	1.848762	1.796387	0.7821277	0.8450980
H1N1	132.0	1.604651	1.507654	0.5758075	0.5880456
H1N1	141.5	1.816746	1.714932	0.7550590	0.7781513
H1N1	165.5	1.850302	1.755659	0.7802756	0.8116246
H3N2	-30.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	-24.0	1.912537	1.870067	0.8138094	0.9030900
H3N2	0.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	2.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	5.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	10.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	12.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	18.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	21.5	1.895526	1.870067	0.8070541	0.9030900
H3N2	26.0	1.941063	1.870067	0.8967332	0.9030900
H3N2	29.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	34.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	36.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	42.0	1.981854	1.903815	0.9284174	0.9286662
H3N2	45.5	1.895526	1.870067	0.8070541	0.9030900
H3N2	50.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	53.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	58.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	60.0	1.823068	1.669139	0.7348510	0.7385606
H3N2	66.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	69.5	1.895526	1.870067	0.8070541	0.9030900

VIRUS	TIMEHOURS	jackson_mean	jackson_med	log_mean	log_med
H3N2	74.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	77.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	82.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	84.0	1.847337	1.792499	0.7690592	0.8406206
H3N2	90.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	93.5	1.895526	1.870067	0.8070541	0.9030900
H3N2	98.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	101.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	106.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	108.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	114.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	117.5	1.895526	1.870067	0.8070541	0.9030900
H3N2	122.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	125.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	130.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	132.0	1.895526	1.870067	0.8070541	0.9030900
H3N2	138.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	141.5	1.845432	1.792499	0.7680348	0.8406206
H3N2	146.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	162.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	165.5	1.895526	1.870067	0.8070541	0.9030900
H3N2	170.0	1.954276	1.870067	0.9107676	0.9030900
H3N2	672.0	1.954276	1.870067	0.9107676	0.9030900
Rhinovirus	-24.0	2.018561	2.058214	0.9471071	1.0413927
Rhinovirus	0.0	1.974378	2.029107	0.9120633	1.0206963
Rhinovirus	4.0	1.996878	2.058214	0.9274206	1.0413927
Rhinovirus	8.0	2.003051	2.058214	0.9379922	1.0413927
Rhinovirus	12.0	2.005789	2.058214	0.9354519	1.0413927
Rhinovirus	16.0	1.989649	2.058214	0.9235932	1.0413927
Rhinovirus	20.0	1.989649	2.058214	0.9235932	1.0413927
Rhinovirus	24.0	1.980491	2.029107	0.9179776	1.0206963
Rhinovirus	30.0	1.996878	2.058214	0.9274206	1.0413927
Rhinovirus	36.0	1.973686	2.058214	0.9096867	1.0413927
Rhinovirus	42.0	1.986407	2.029107	0.9194988	1.0206963
Rhinovirus	48.0	1.974378	2.029107	0.9120633	1.0206963
Rhinovirus	60.0	2.097366	2.058214	1.0031317	1.0413927
Rhinovirus	72.0	1.983866	2.058214	0.9178226	1.0413927
Rhinovirus	84.0	2.157074	2.346400	1.0358571	1.2304489
Rhinovirus	96.0	1.968556	2.058214	0.9045007	1.0413927
Rhinovirus	108.0	2.192423	2.535768	1.0501051	1.3424227
Rhinovirus	120.0	1.987402	2.029107	0.9146618	1.0206963
Rhinovirus	136.0	2.158317	2.441084	1.0306798	1.2864358
RSV	-24.0	1.952080	1.693975	0.8690260	0.7525750
RSV	0.0	1.952080	1.693975	0.8690260	0.7525750
RSV	5.0	1.952080	1.693975	0.8690260	0.7525750
RSV	12.0	1.952080	1.693975	0.8690260	0.7525750
RSV	21.5	1.952080	1.693975	0.8690260	0.7525750
RSV	29.0	1.943619	1.517882	0.8579652	0.6020600
RSV	36.0	1.952080	1.693975	0.8690260	0.7525750
RSV	45.5	1.952080	1.693975	0.8690260	0.7525750
RSV	53.0	1.909786	1.517882	0.8377959	0.6020600
RSV	60.0	1.952080	1.693975	0.8690260	0.7525750

VIRUS	TIMEHOURS	jackson_mean	jackson_med	log_mean	log_med
RSV	69.5	1.952080	1.693975	0.8690260	0.7525750
RSV	77.0	1.872511	1.517882	0.8204278	0.6020600
RSV	84.0	1.952080	1.693975	0.8690260	0.7525750
RSV	93.5	1.952080	1.693975	0.8690260	0.7525750
RSV	101.0	1.952080	1.693975	0.8690260	0.7525750
RSV	108.0	1.952080	1.693975	0.8690260	0.7525750
RSV	117.5	1.952080	1.693975	0.8690260	0.7525750
RSV	125.0	1.952080	1.693975	0.8690260	0.7525750
RSV	132.0	1.952080	1.693975	0.8690260	0.7525750
RSV	141.5	1.952080	1.693975	0.8690260	0.7525750
RSV	165.5	1.952080	1.693975	0.8690260	0.7525750

```
z <- train_symptom_df %>%
  dplyr::group_by(STUDYID) %>%
  dplyr::rowwise() %>%
  dplyr::mutate(SUM_SCORE = sum(SX_HEADACHE,
                                SX_RUNNYNOSE,
                                SX_COUGH,
                                SX_MALAISE,
                                SX_MYALGIA,
                                SX_SNEEZE,
                                SX_SORETHROAT,
                                SX_STUFFYNOSE))
z %>% dplyr::ungroup() %>% dplyr::group_by(STUDYID) %>%
  dplyr::summarise(score_mean=mean(SUM_SCORE),
                   score_med=median(SUM_SCORE)) %>%
  pander(split.cells = 50, split.table = Inf)
```

STUDYID	score_mean	score_med
DEE1 RSV	1.300000	0
DEE2 H3N2	1.647059	0
DEE3 H1N1	1.279167	0
DEE4X H1N1	1.025000	0
DEE5 H3N2	1.435897	0
Rhinovirus Duke	2.142857	1
Rhinovirus UVA	1.435714	1

```
z <- train_symptom_df %>%
  dplyr::mutate(STUDYDAY = as.factor(STUDYDAY)) %>%
  dplyr::group_by(STUDYID,STUDYDAY) %>%
  dplyr::rowwise() %>%
  dplyr::mutate(SUM_SCORE = sum(SX_HEADACHE,
                                SX_RUNNYNOSE,
                                SX_COUGH,
                                SX_MALAISE,
                                SX_MYALGIA,
                                SX_SNEEZE,
                                SX_SORETHROAT,
                                SX_STUFFYNOSE))
```

```
z %>% dplyr::ungroup() %>% dplyr::group_by(STUDYID,STUDYDAY) %>%
  dplyr::summarise(score_mean=mean(SUM_SCORE),
                   score_med=median(SUM_SCORE)) %>%
  pander(split.cells = 50, split.table = Inf)
```

STUDYID	STUDYDAY	score_mean	score_med
DEE1 RSV	-2	0.07692308	0.0
DEE1 RSV	-1	0.00000000	0.0
DEE1 RSV	0	0.23076923	0.0
DEE1 RSV	1	0.15384615	0.0
DEE1 RSV	2	0.46153846	0.0
DEE1 RSV	3	1.15384615	0.0
DEE1 RSV	4	2.30769231	1.0
DEE1 RSV	5	2.53846154	2.0
DEE1 RSV	6	3.38461538	1.0
DEE1 RSV	7	2.69230769	1.0
DEE2 H3N2	0	0.29411765	0.0
DEE2 H3N2	1	1.11764706	0.0
DEE2 H3N2	2	3.35294118	2.0
DEE2 H3N2	3	3.17647059	2.0
DEE2 H3N2	4	2.82352941	2.0
DEE2 H3N2	5	1.64705882	1.0
DEE2 H3N2	6	0.52941176	0.0
DEE2 H3N2	7	0.23529412	0.0
DEE3 H1N1	-2	0.00000000	0.0
DEE3 H1N1	-1	0.08333333	0.0
DEE3 H1N1	0	0.50000000	0.0
DEE3 H1N1	1	1.29166667	0.0
DEE3 H1N1	2	1.58333333	0.0
DEE3 H1N1	3	2.00000000	1.0
DEE3 H1N1	4	3.41666667	2.5
DEE3 H1N1	5	2.45833333	1.0
DEE3 H1N1	6	1.20833333	0.5
DEE3 H1N1	7	0.25000000	0.0
DEE4X H1N1	-2	0.08333333	0.0
DEE4X H1N1	-1	0.25000000	0.0
DEE4X H1N1	0	0.25000000	0.0
DEE4X H1N1	1	0.75000000	1.0
DEE4X H1N1	2	2.08333333	0.5
DEE4X H1N1	3	2.66666667	1.0
DEE4X H1N1	4	1.66666667	0.0
DEE4X H1N1	5	1.50000000	0.5
DEE4X H1N1	6	0.83333333	0.0
DEE4X H1N1	7	0.16666667	0.0
DEE5 H3N2	-1	0.61538462	0.0
DEE5 H3N2	0	0.53846154	0.0
DEE5 H3N2	1	1.61538462	2.0
DEE5 H3N2	2	3.23076923	2.0
DEE5 H3N2	3	2.84615385	1.0
DEE5 H3N2	4	1.46153846	1.0
DEE5 H3N2	5	1.53846154	1.0
DEE5 H3N2	6	0.53846154	0.0
DEE5 H3N2	7	0.53846154	0.0

STUDYID	STUDYDAY	score_mean	score_med
Rhinovirus Duke	-1	0.10526316	0.0
Rhinovirus Duke	0	0.73684211	1.0
Rhinovirus Duke	1	1.36842105	1.0
Rhinovirus Duke	2	2.63157895	2.0
Rhinovirus Duke	3	3.36842105	3.0
Rhinovirus Duke	4	3.57894737	2.0
Rhinovirus Duke	5	3.21052632	2.0
Rhinovirus UVA	-1	0.00000000	0.0
Rhinovirus UVA	0	0.45000000	0.0
Rhinovirus UVA	1	1.50000000	1.0
Rhinovirus UVA	2	1.55000000	1.0
Rhinovirus UVA	3	2.35000000	1.5
Rhinovirus UVA	4	2.20000000	1.5
Rhinovirus UVA	5	2.00000000	1.0

```r
z <- train_symptom_df %>%
  add_virus_col() %>%
  dplyr::group_by(VIRUS) %>%
  dplyr::rowwise() %>%
  dplyr::mutate(SUM_SCORE = sum(SX_HEADACHE,
                               SX_RUNNYNOSE,
                               SX_COUGH,
                               SX_MALAISE,
                               SX_MYALGIA,
                               SX_SNEEZE,
                               SX_SORETHROAT,
                               SX_STUFFYNOSE))
z %>% dplyr::ungroup() %>% dplyr::group_by(VIRUS) %>%
  dplyr::summarise(score_mean=mean(SUM_SCORE),
                  score_med=median(SUM_SCORE)) %>%
  pander(split.cells = 50, split.table = Inf)
```

VIRUS	score_mean	score_med
H1N1	1.194444	0
H3N2	1.549407	0
Rhinovirus	1.780220	1
RSV	1.300000	0

```r
z<- train_symptom_df %>%
  add_virus_col() %>%
  dplyr::mutate(STUDYDAY = as.factor(STUDYDAY)) %>%
  dplyr::group_by(VIRUS,STUDYDAY) %>%
  dplyr::rowwise() %>%
  dplyr::mutate(SUM_SCORE = sum(SX_HEADACHE,
                               SX_RUNNYNOSE,
                               SX_COUGH,
                               SX_MALAISE,
                               SX_MYALGIA,
                               SX_SNEEZE,
```

```
                                SX_SORETHROAT,
                                SX_STUFFYNOSE))
z %>% dplyr::ungroup() %>% dplyr::group_by(VIRUS,STUDYDAY) %>%
  dplyr::summarise(score_mean=mean(SUM_SCORE),
                score_med=median(SUM_SCORE)) %>%
  pander(split.cells = 50, split.table = Inf)
```

VIRUS	STUDYDAY	score_mean	score_med
H1N1	-2	0.02777778	0.0
H1N1	-1	0.13888889	0.0
H1N1	0	0.41666667	0.0
H1N1	1	1.11111111	0.0
H1N1	2	1.75000000	0.0
H1N1	3	2.22222222	1.0
H1N1	4	2.83333333	1.5
H1N1	5	2.13888889	1.0
H1N1	6	1.08333333	0.0
H1N1	7	0.22222222	0.0
H3N2	-1	0.61538462	0.0
H3N2	0	0.40000000	0.0
H3N2	1	1.33333333	0.0
H3N2	2	3.30000000	2.0
H3N2	3	3.03333333	1.5
H3N2	4	2.23333333	1.0
H3N2	5	1.60000000	1.0
H3N2	6	0.53333333	0.0
H3N2	7	0.36666667	0.0
Rhinovirus	-1	0.05128205	0.0
Rhinovirus	0	0.58974359	0.0
Rhinovirus	1	1.43589744	1.0
Rhinovirus	2	2.07692308	2.0
Rhinovirus	3	2.84615385	2.0
Rhinovirus	4	2.87179487	2.0
Rhinovirus	5	2.58974359	1.0
RSV	-2	0.07692308	0.0
RSV	-1	0.00000000	0.0
RSV	0	0.23076923	0.0
RSV	1	0.15384615	0.0
RSV	2	0.46153846	0.0
RSV	3	1.15384615	0.0
RSV	4	2.30769231	1.0
RSV	5	2.53846154	2.0
RSV	6	3.38461538	1.0
RSV	7	2.69230769	1.0

Patient Log Symptom Score Distribution by Study ID:
Train Data
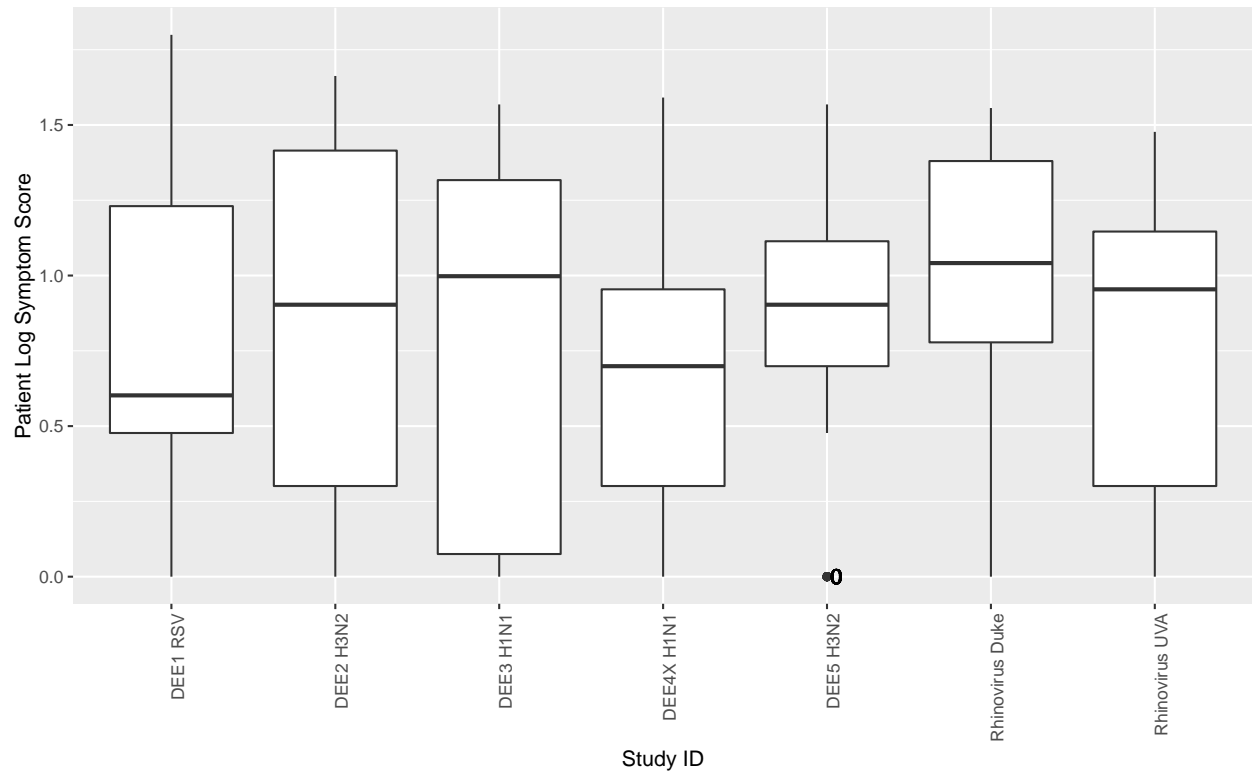
Patient Log Symptom Score Distribution by Virus:
Train Data
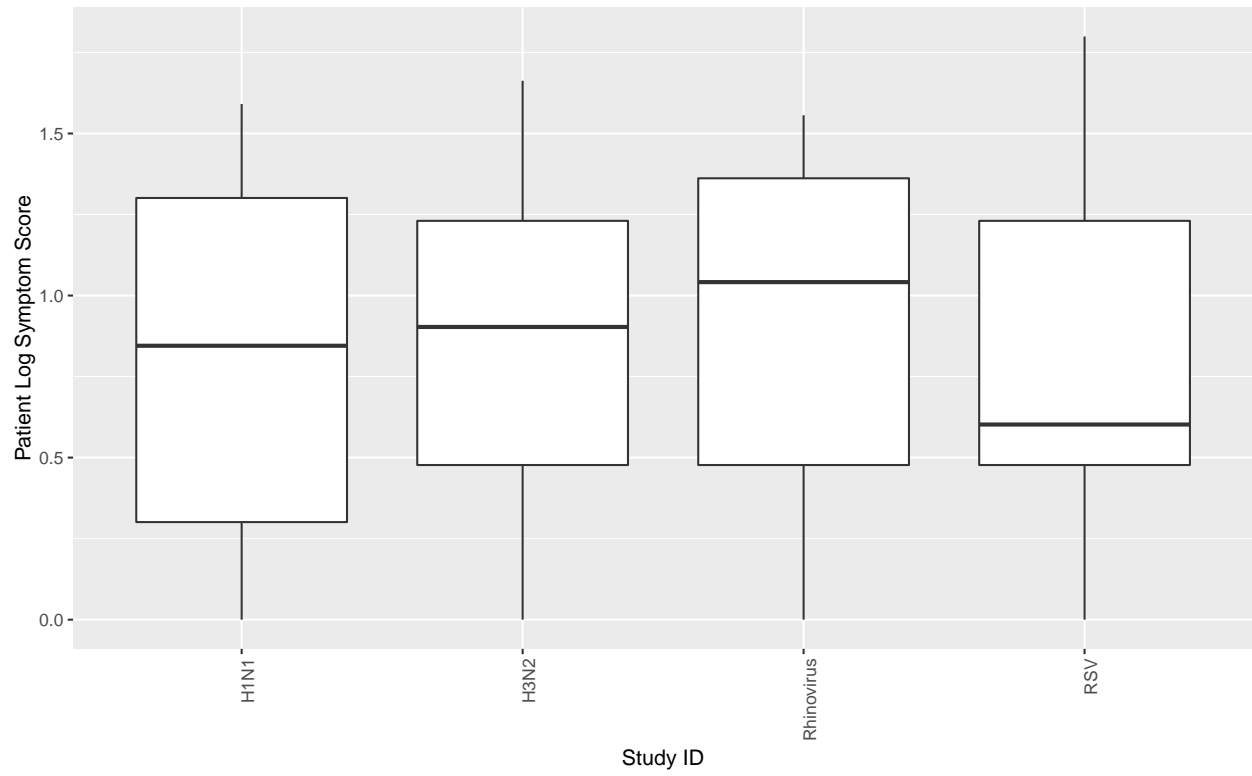
Patient Jackson Score Distribution by Study ID:
Train Data
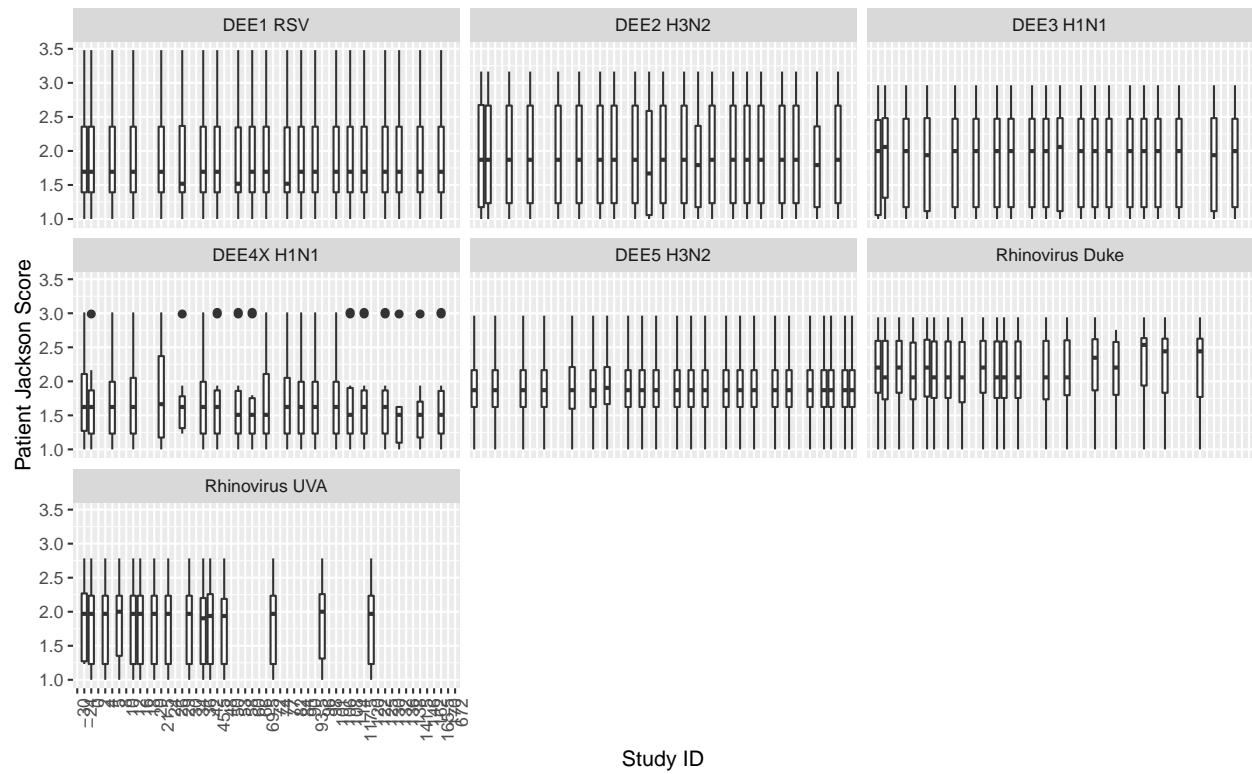
Patient Jackson Score Distribution by Virus:
Train Data
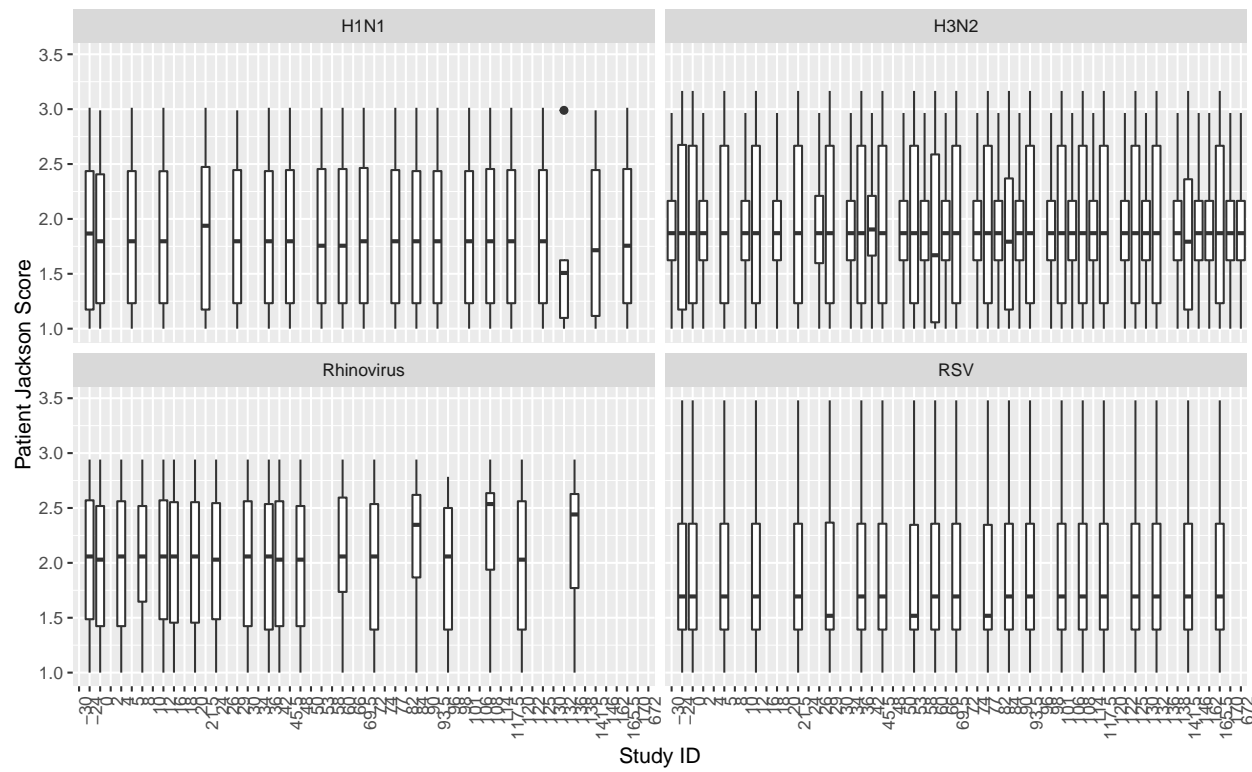
Patient Symptom Score Sum Distribution by Study ID:
Train Data
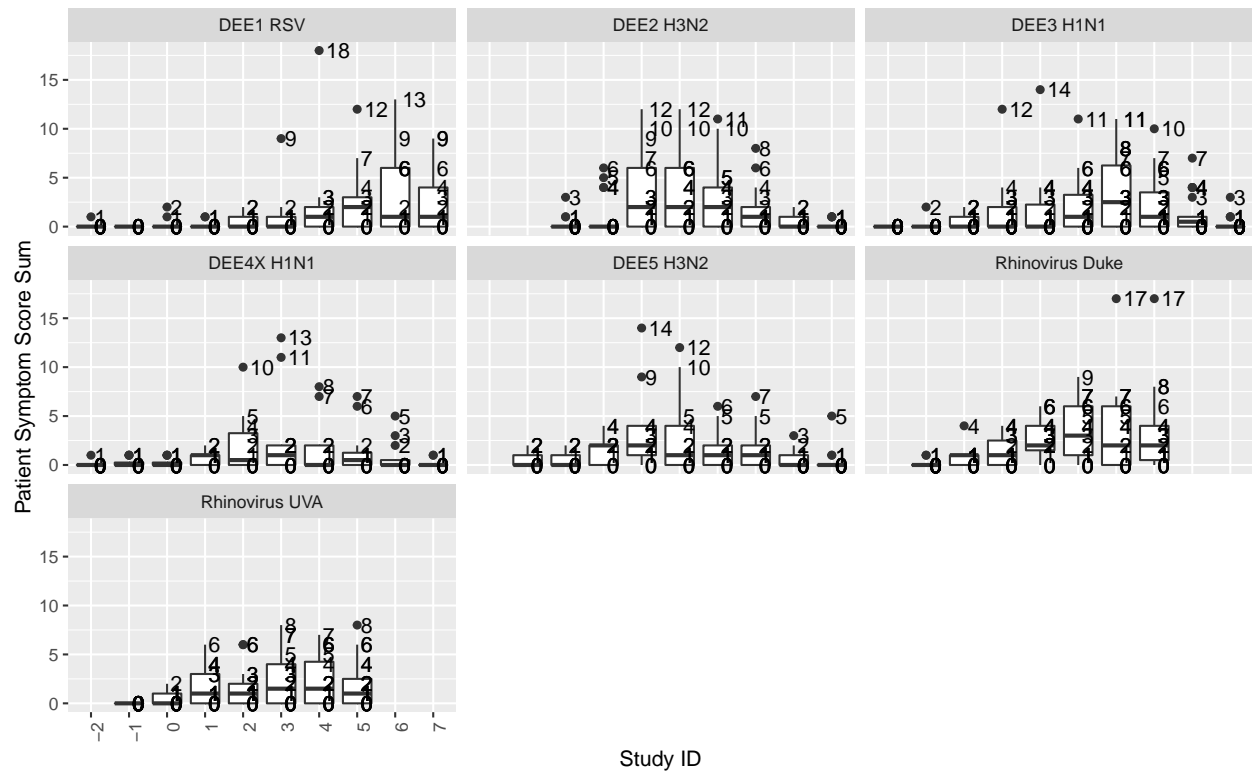
Patient Symptom Score Sum Distribution by Virus:
Train Data
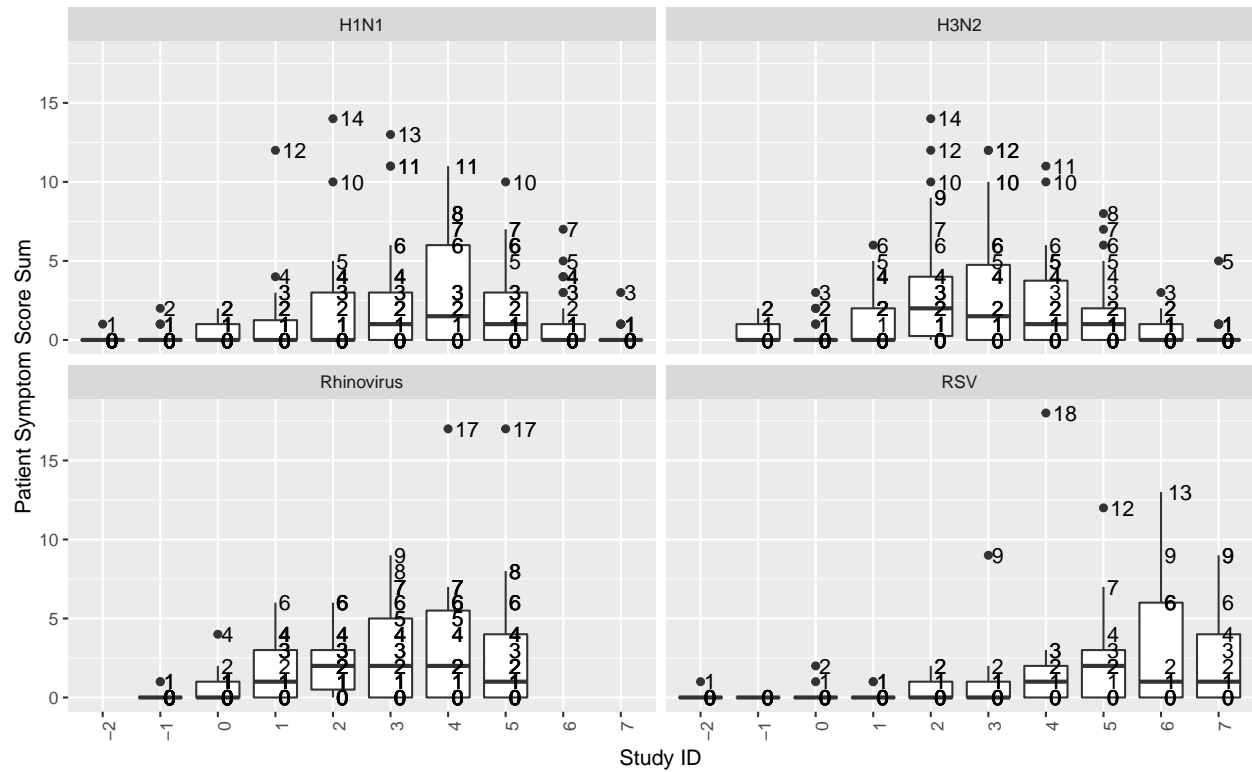
Ensure all patients are between 15-60 yo

```
train_clinicl_df$AGE %>% summary()
```

Min. 1st Qu. Median Mean 3rd Qu. Max. 18.00 21.00 24.00 25.19 27.00 45.00
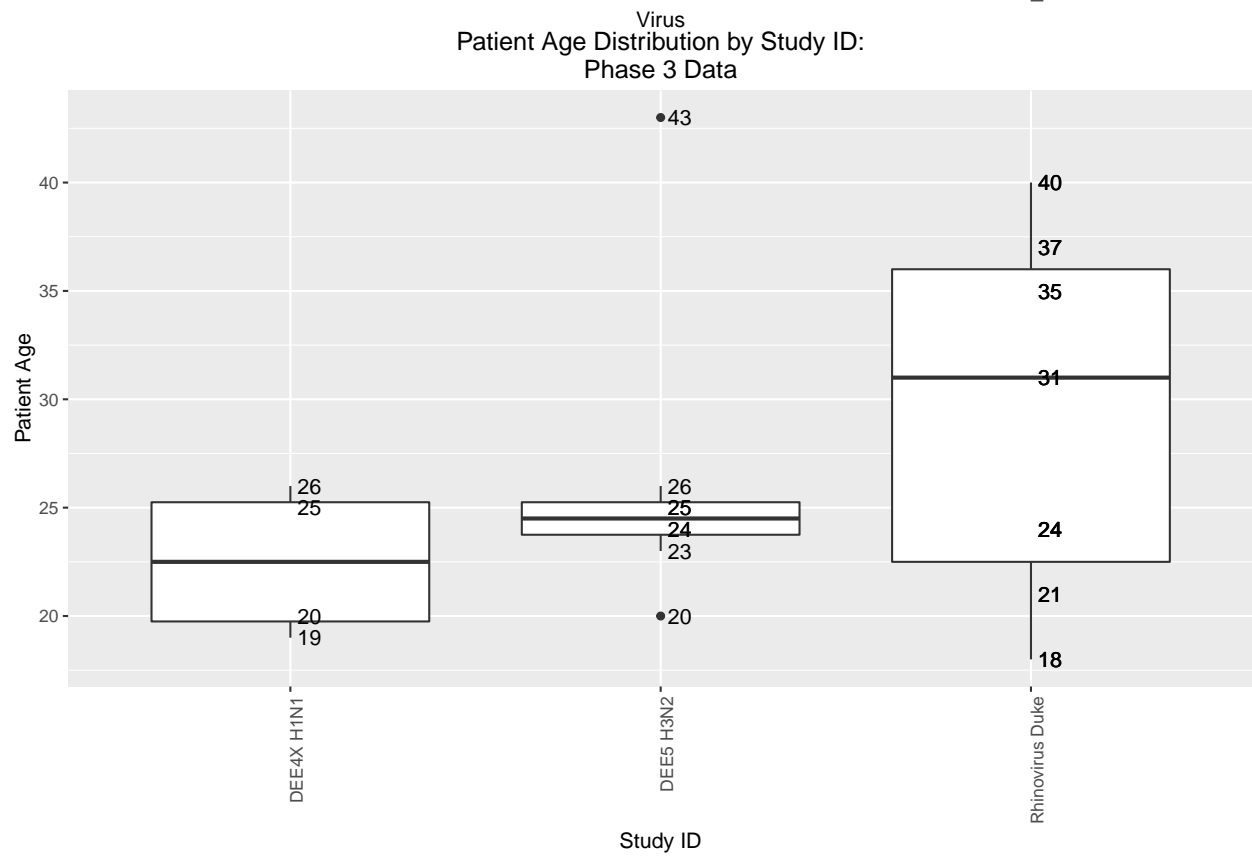
```
phas1_clinicl_df$AGE %>% summary()
```

Min. 1st Qu. Median Mean 3rd Qu. Max. 18.00 21.25 24.00 26.15 29.00 43.00

```
phas2_clinicl_df$AGE %>% summary()
```

Min. 1st Qu. Median Mean 3rd Qu. Max. 18.00 22.00 24.00 26.31 30.25 43.00
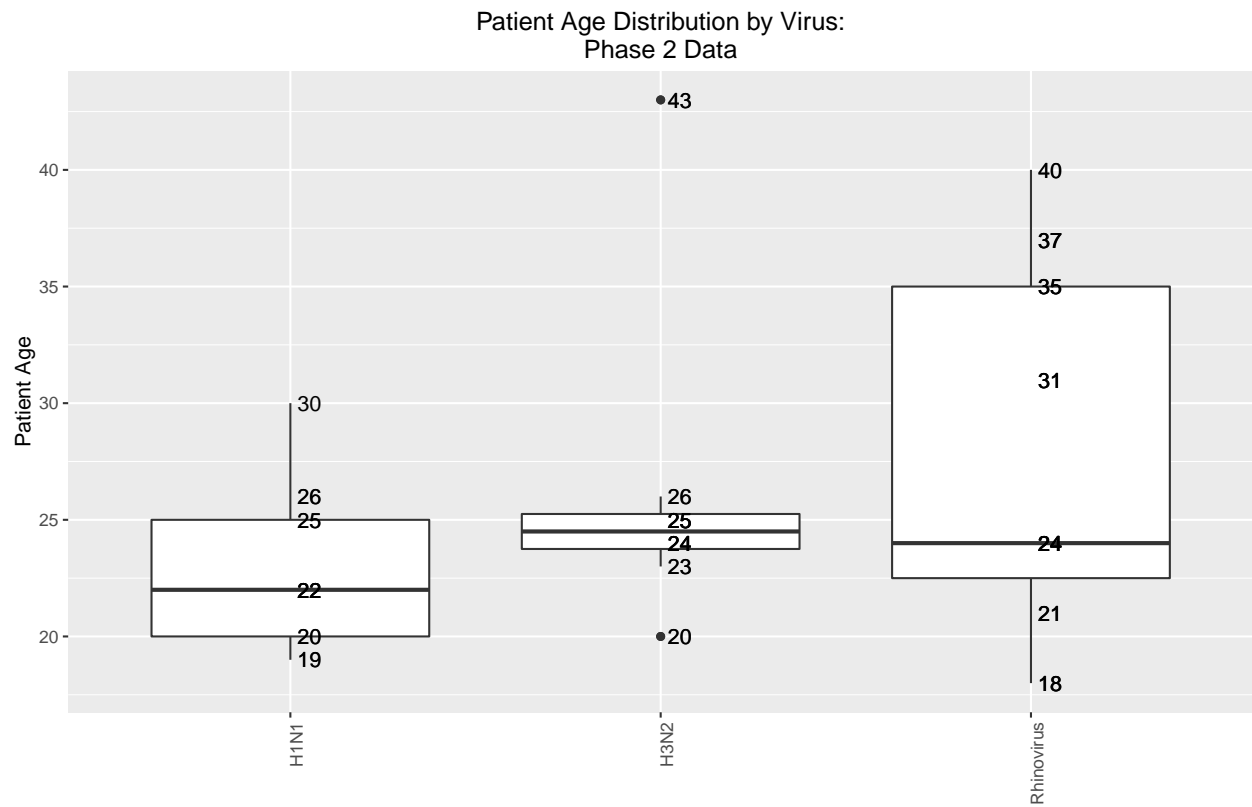
```
phas3_clinicl_df$AGE %>% summary()
```

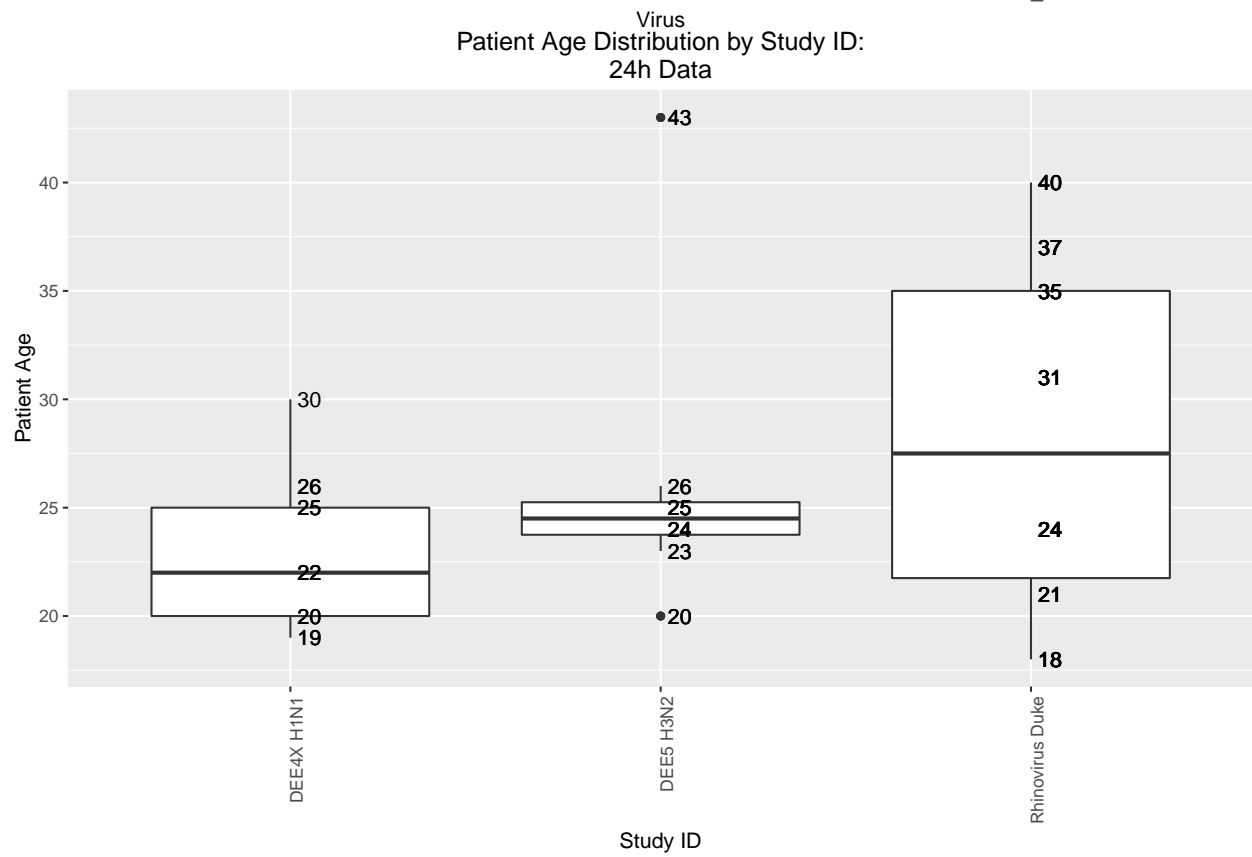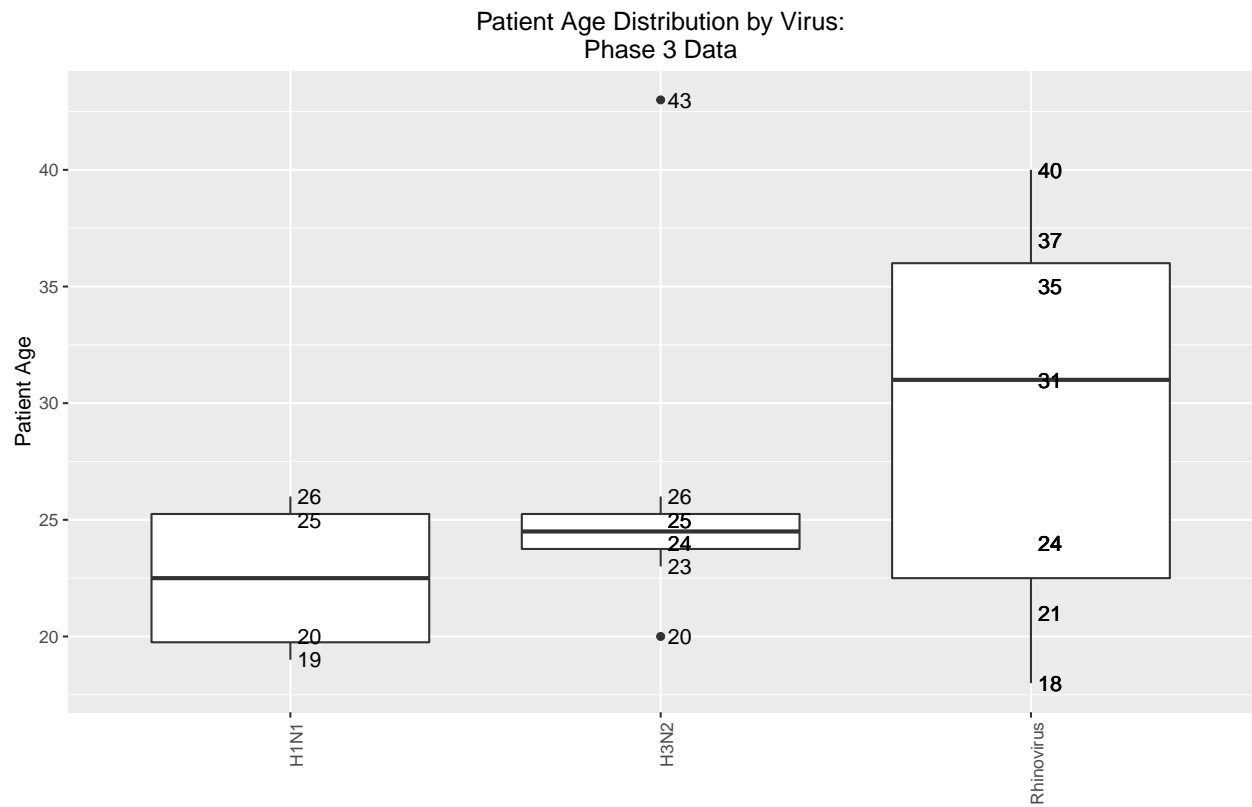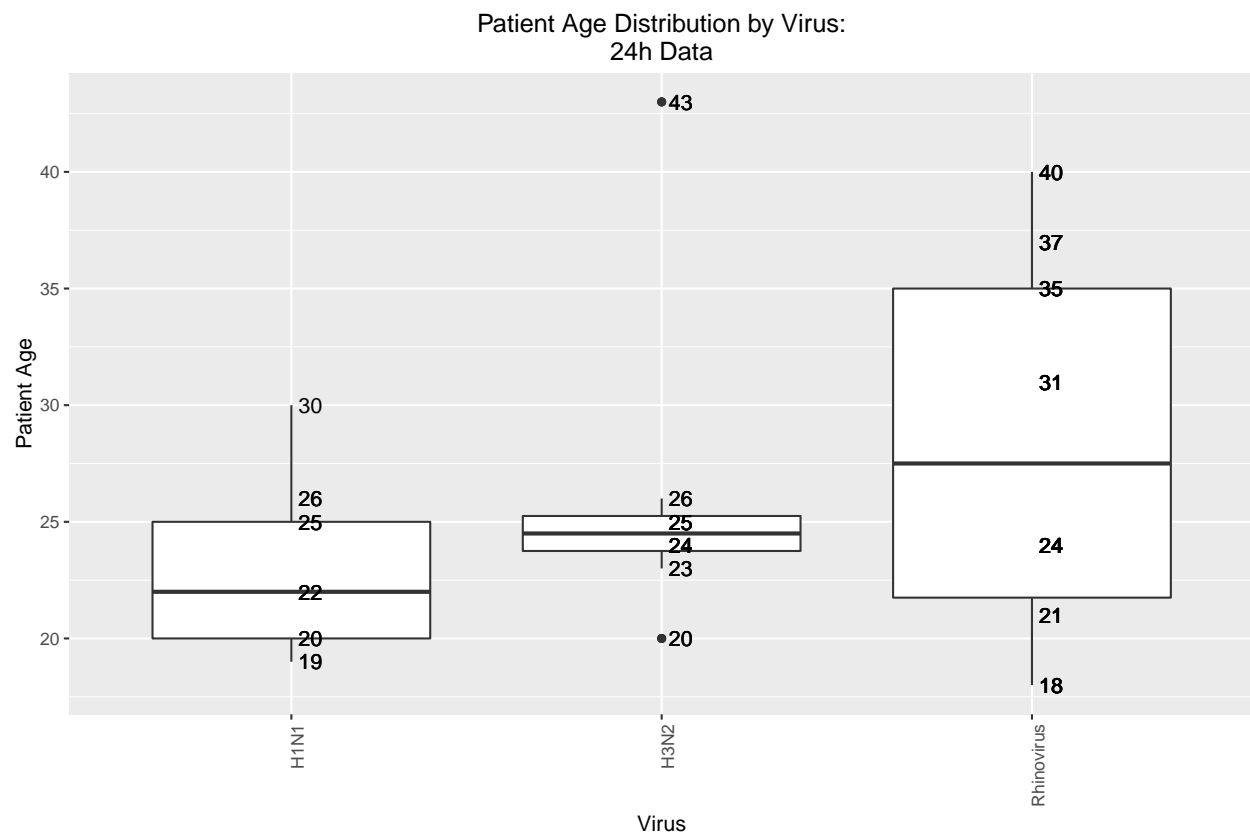Min. 1st Qu. Median Mean 3rd Qu. Max. 18.0 22.0 25.0 27.6 35.0 43.0

Patient Age Distribution by Study ID: Train Data

Patient Age Distribution by Virus:
Train Data

Patient Age Distribution by Study ID:
Phase 1 (0h) Data

Patient Age Distribution by Virus:
Phase 1 (0h) Data

Patient Age Distribution by Study ID:
Phase 2 Data

Patient Age Distribution by Virus:
Phase 2 Data

Patient Age Distribution by Study ID:
Phase 3 Data

Patient Age Distribution by Virus:
Phase 3 Data

Patient Age Distribution by Virus: Phase 3 Data — boxplot with Patient Age on y-axis and Virus (H1N1, H3N2, Rhinovirus) on x-axis.

Patient Age Distribution by Study ID:
24h Data

Patient Age Distribution by Study ID: 24h Data — boxplot with Patient Age on y-axis and Study ID (DEE4X H1N1, DEE5 H3N2, Rhinovirus Duke) on x-axis.

Patient Age Distribution by Virus:
24h Data

[Figure: Box plot showing Patient Age distribution by Virus (H1N1, H3N2, Rhinovirus). Y-axis: Patient Age (20, 25, 30, 35, 40). X-axis: Virus. H1N1 box ranges from 20 to 25 with median 22, whisker to 30, labels 19, 20, 22, 25, 26, 30. H3N2 box ranges from 24 to 25 with median 24.5, labels 23, 24, 25, 26, outliers at 20 and 43. Rhinovirus box ranges from 21 to 35 with median ~27.5, labels 18, 21, 24, 31, 35, 37, 40.]

Look at removing patients who report pre-challenge symptoms

```
train_clinicl_df %>%
  dplyr::distinct(STUDYID,SUBJECTID) %>% dim()
```

[1] 125 2

```
train_clinicl_df %>%
  dplyr::filter(TIMEHOURS <= 0, LOGSYMPTSCORE_SC3 > 0) %>%
  dplyr::distinct(STUDYID,SUBJECTID) %>% dim()
```

[1] 107 2

```
train_symptom_df %>%
  dplyr::distinct(STUDYID,SUBJECTID) %>% dim()
```

[1] 118 2

```
train_symptom_df %>%
  dplyr::rowwise() %>%
  dplyr::mutate(SUM_SCORE = sum(SX_HEADACHE,
                               SX_RUNNYNOSE,
                               SX_COUGH,
                               SX_MALAISE,
```

```
                                SX_MYALGIA,
                                SX_SNEEZE,
                                SX_SORETHROAT,
                                SX_STUFFYNOSE)) %>%
  dplyr::filter(STUDYDAY <= 0,SUM_SCORE > 0) %>%
  dplyr::distinct(STUDYID,SUBJECTID) %>% dim()
```
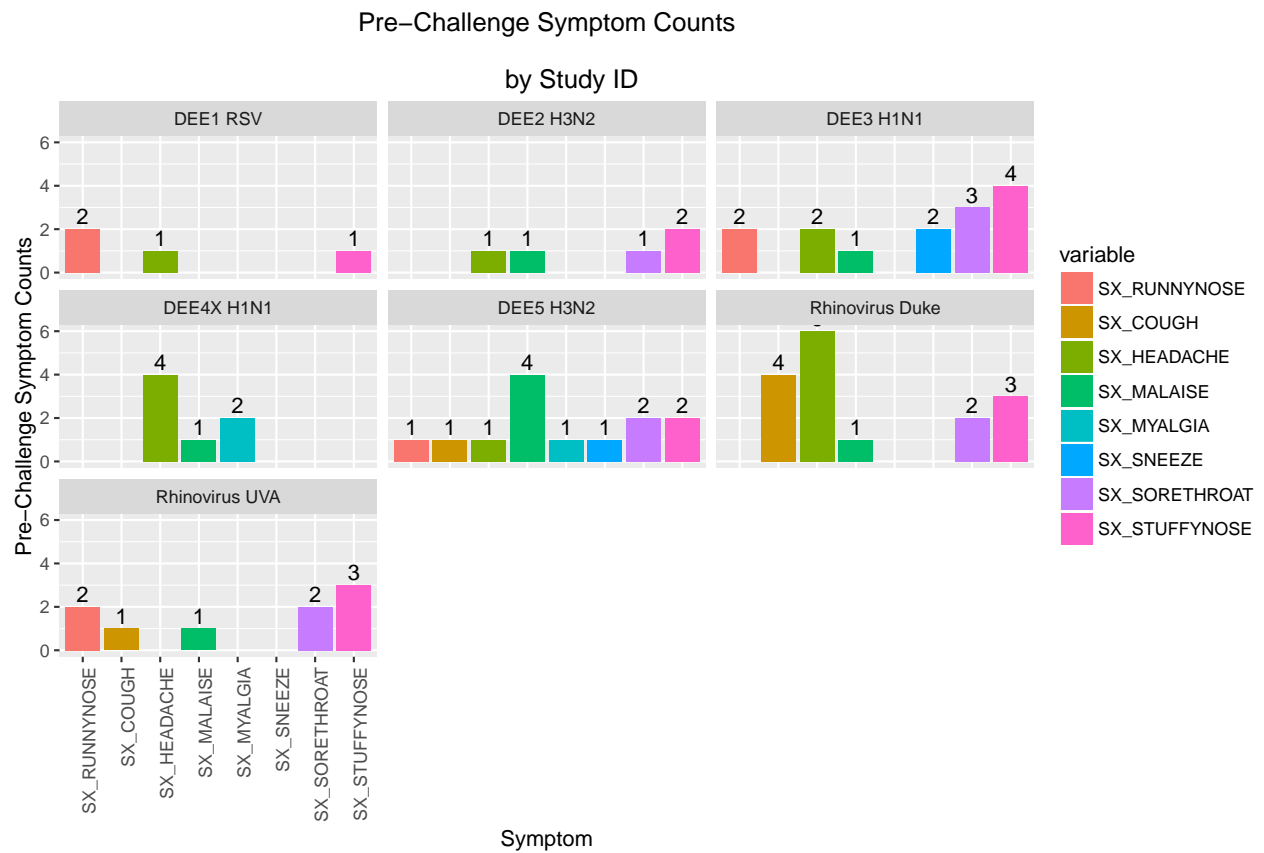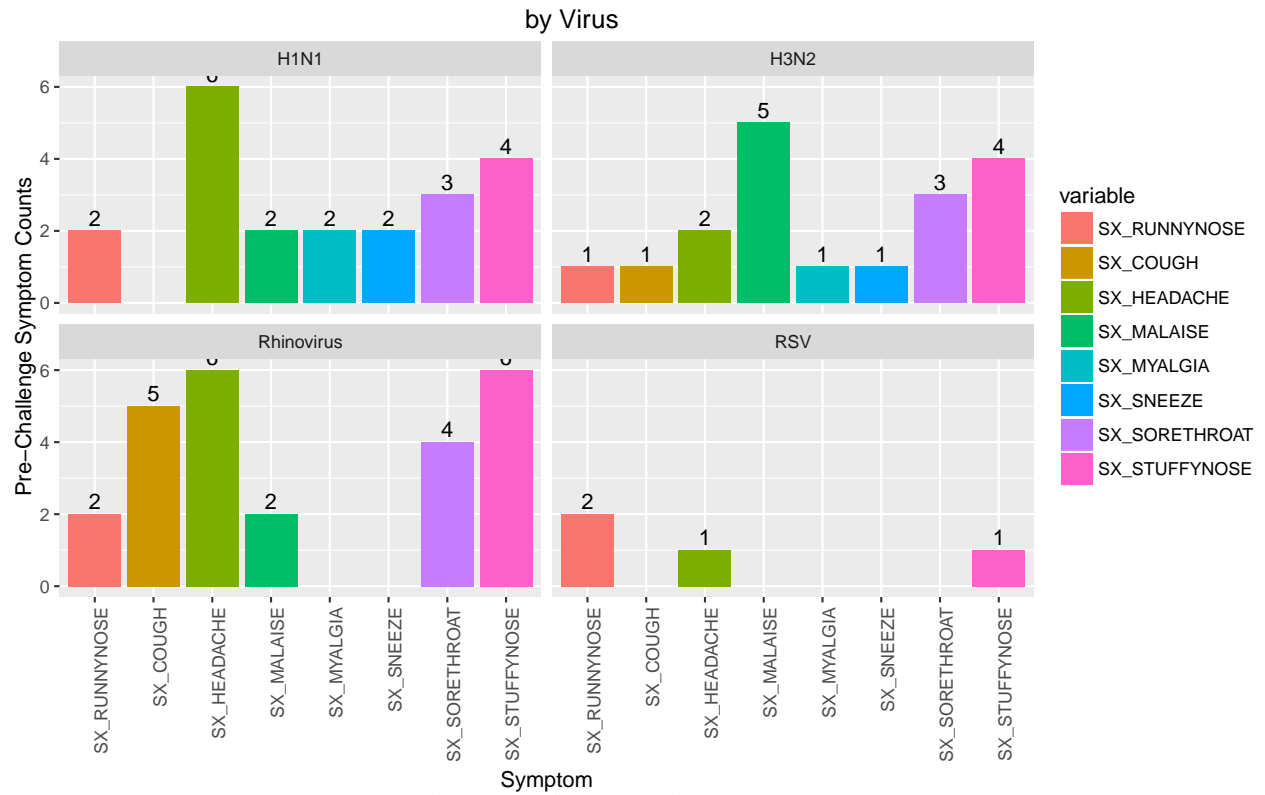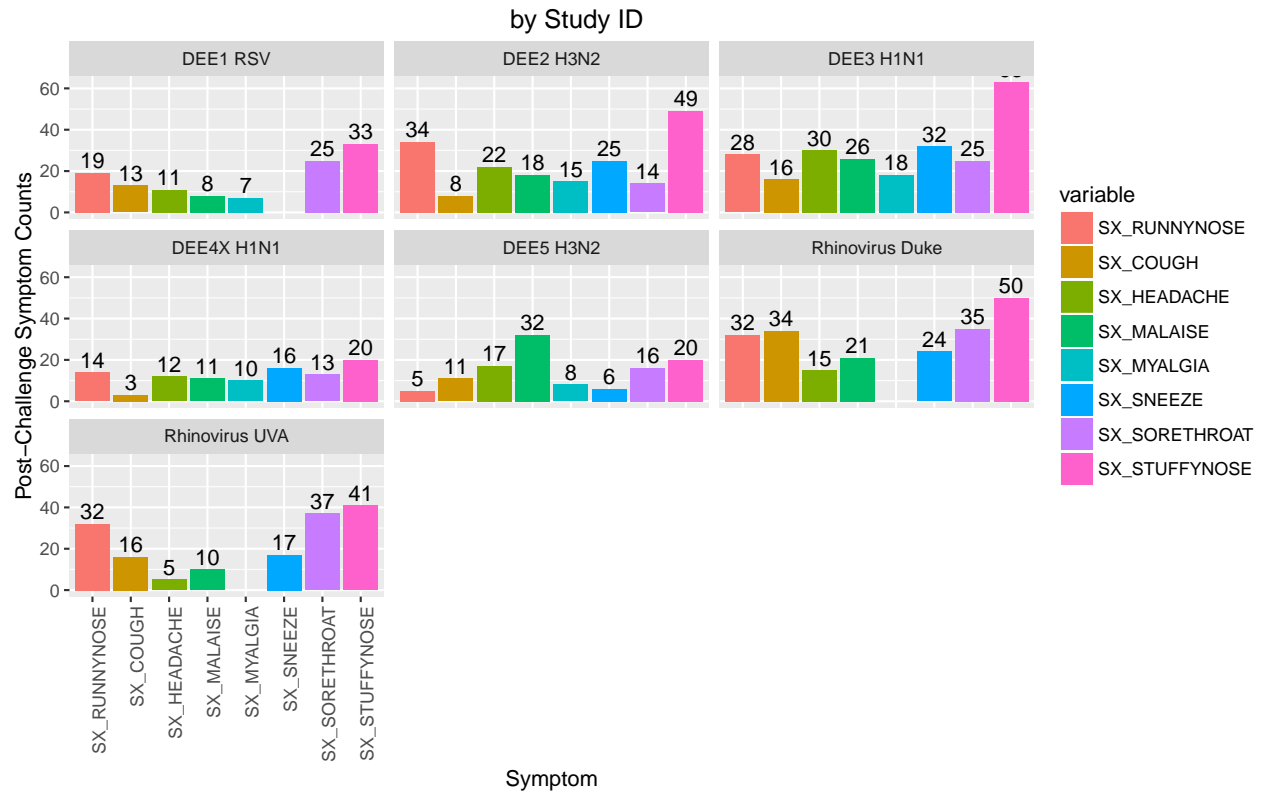
[1] 47 2

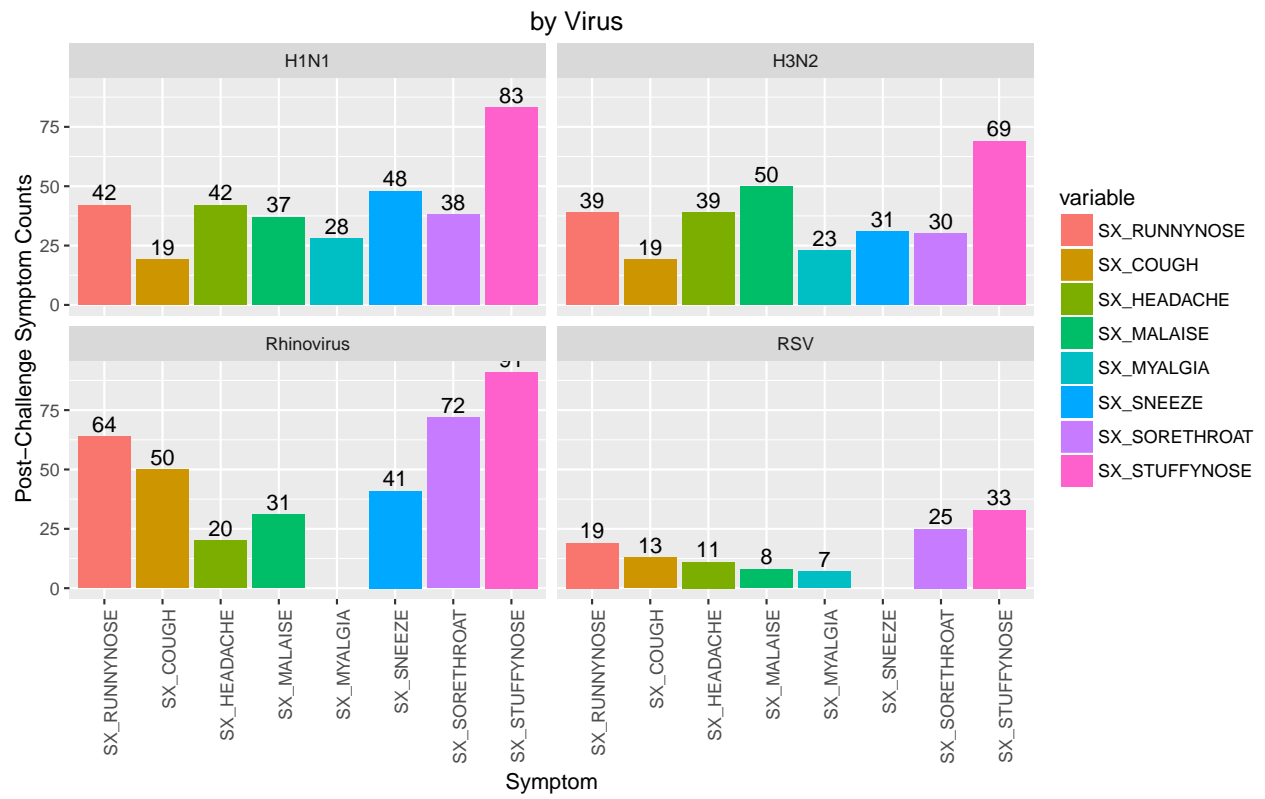**Could we use pre-challenge unreported symptoms as predictors?**
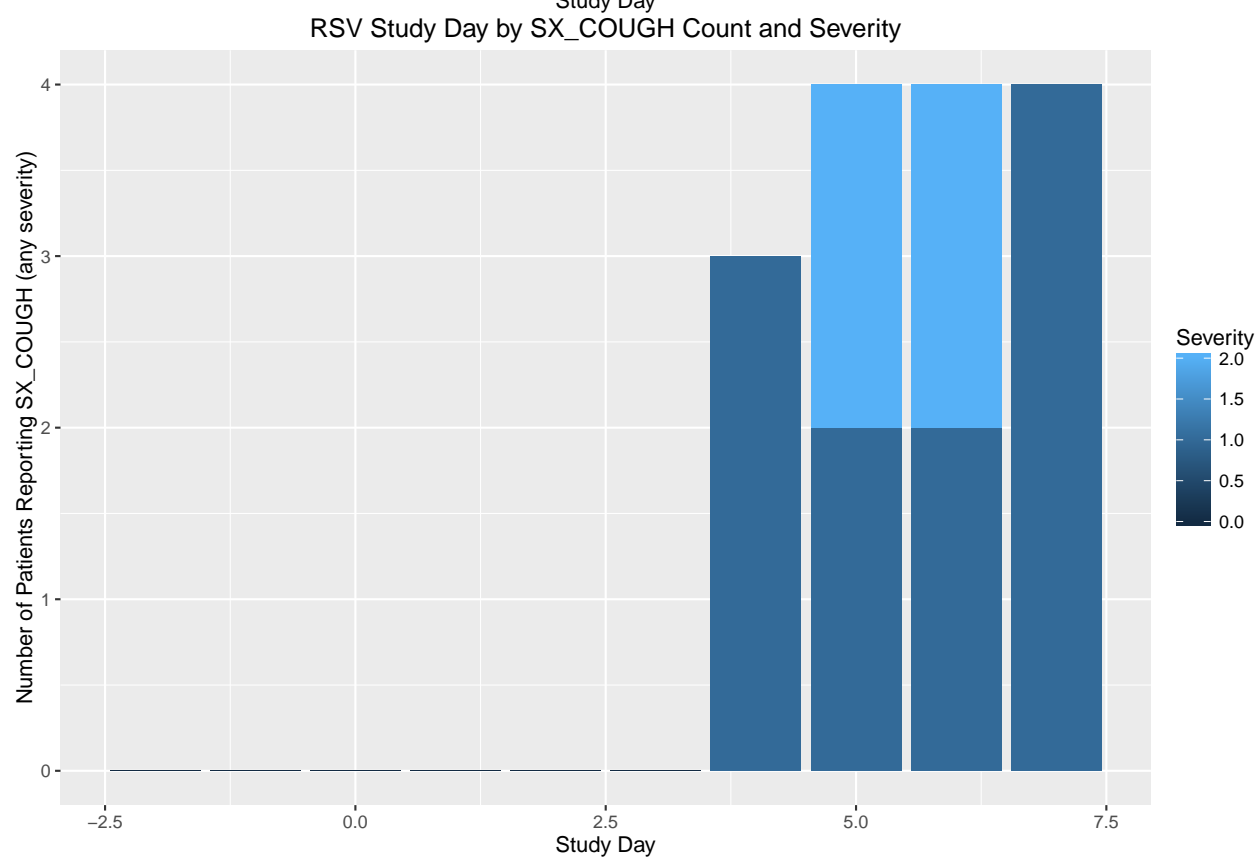

Pre–Challenge Symptom Counts

by Study ID

Pre−Challenge Symptom Counts

by Virus

Post−Challenge Symptom Counts
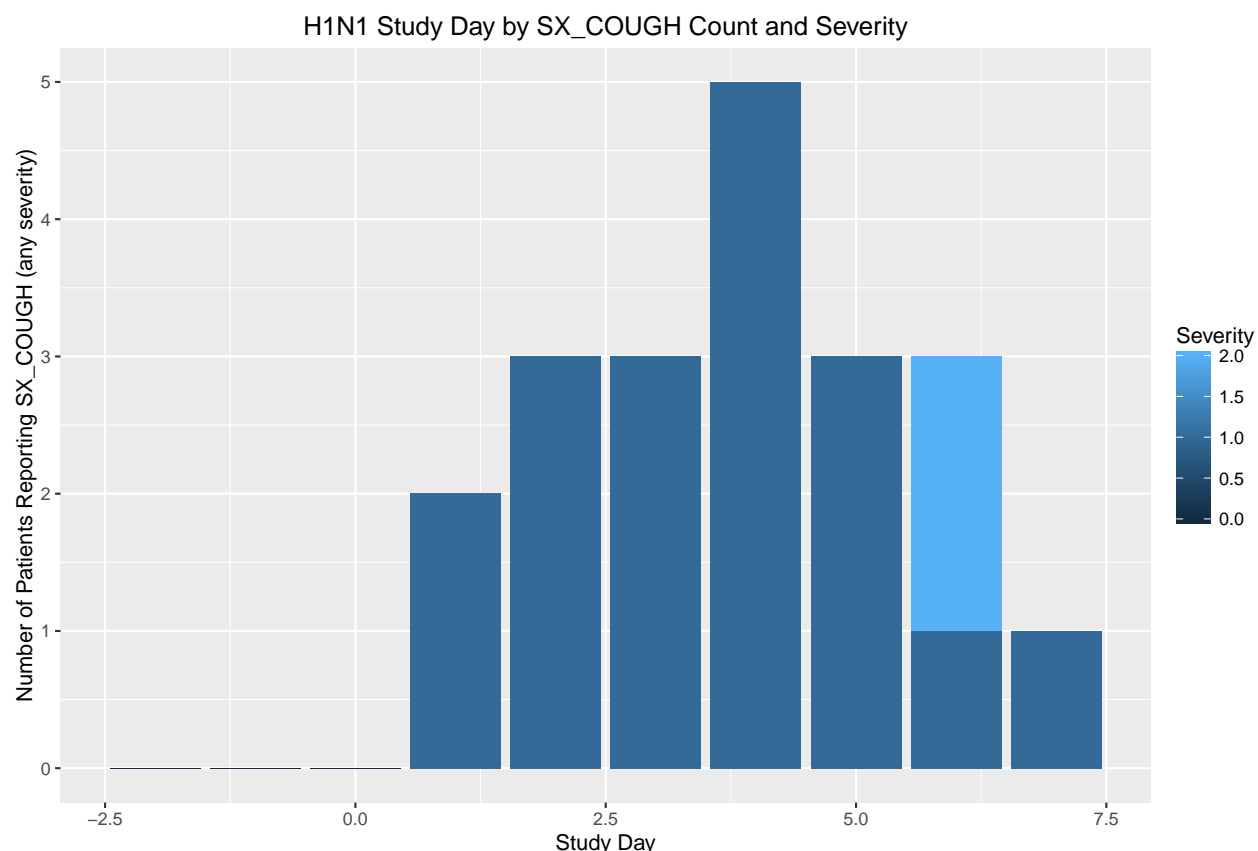
by Study ID

Post−Challenge Symptom Counts

by Virus

What are the distributions of pre-challenge unreported symptoms?

Rhinovirus Study Day by SX_SNEEZE Count and Severity

H1N1 Study Day by SX_COUGH Count and Severity

RSV Study Day by SX_COUGH Count and Severity

RSV Study Day by SX_MALAISE Count and Severity

RSV Study Day by SX_MYALGIA Count and Severity

RSV Study Day by SX_SORETHROAT Count and Severity
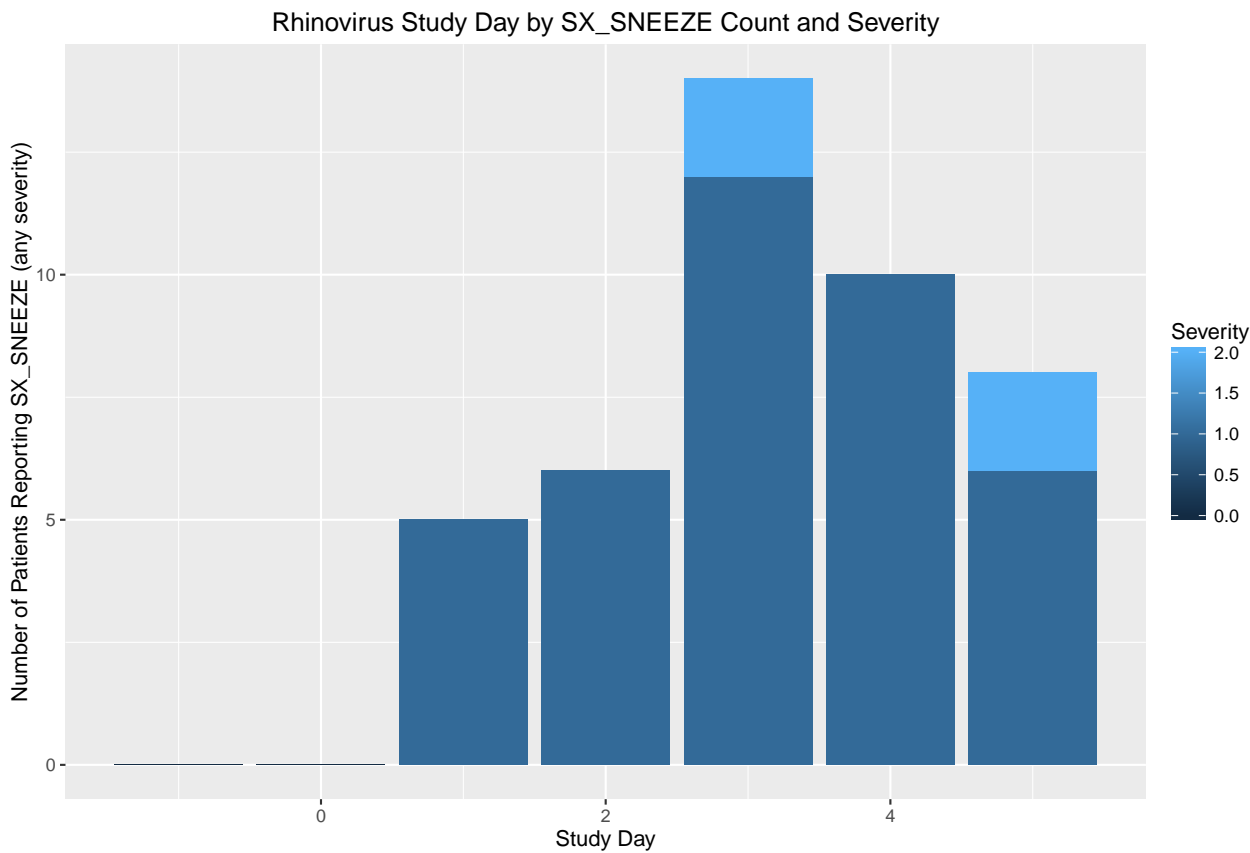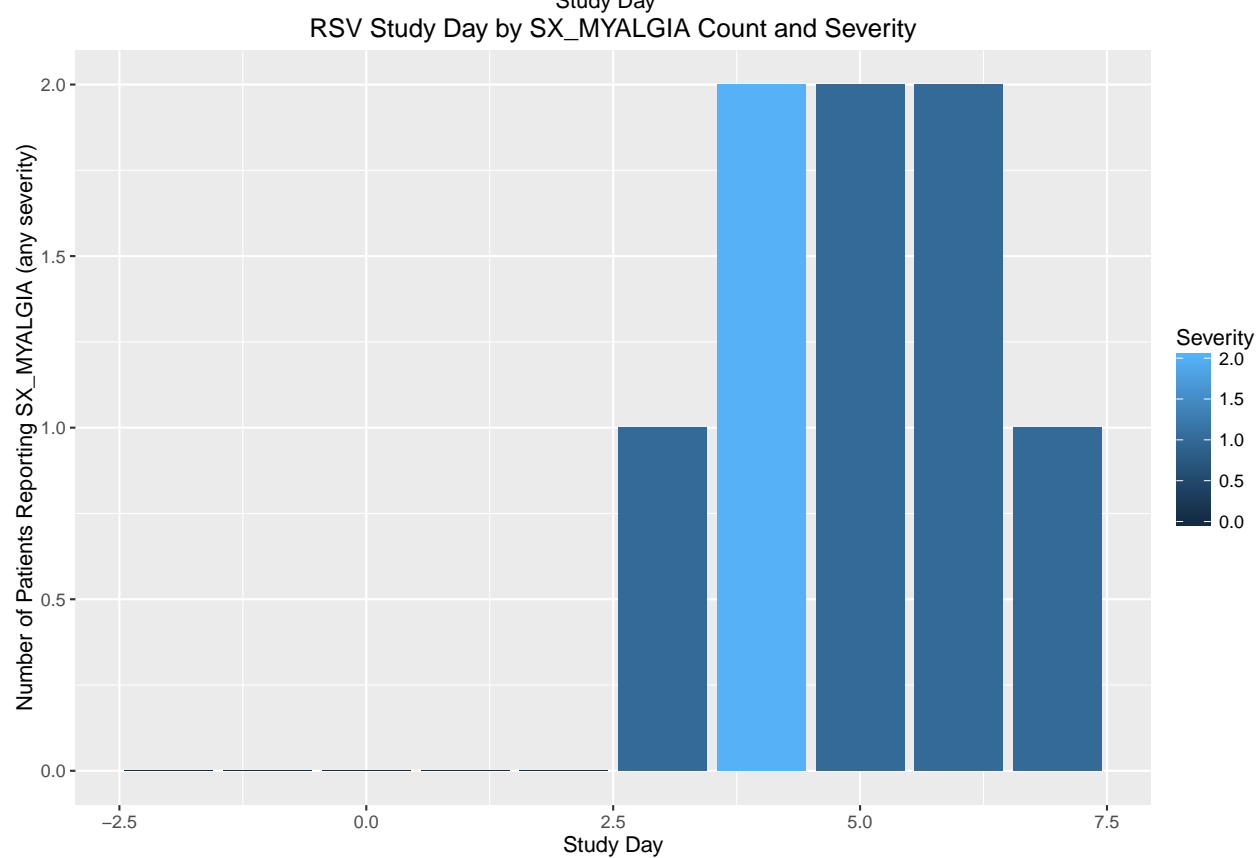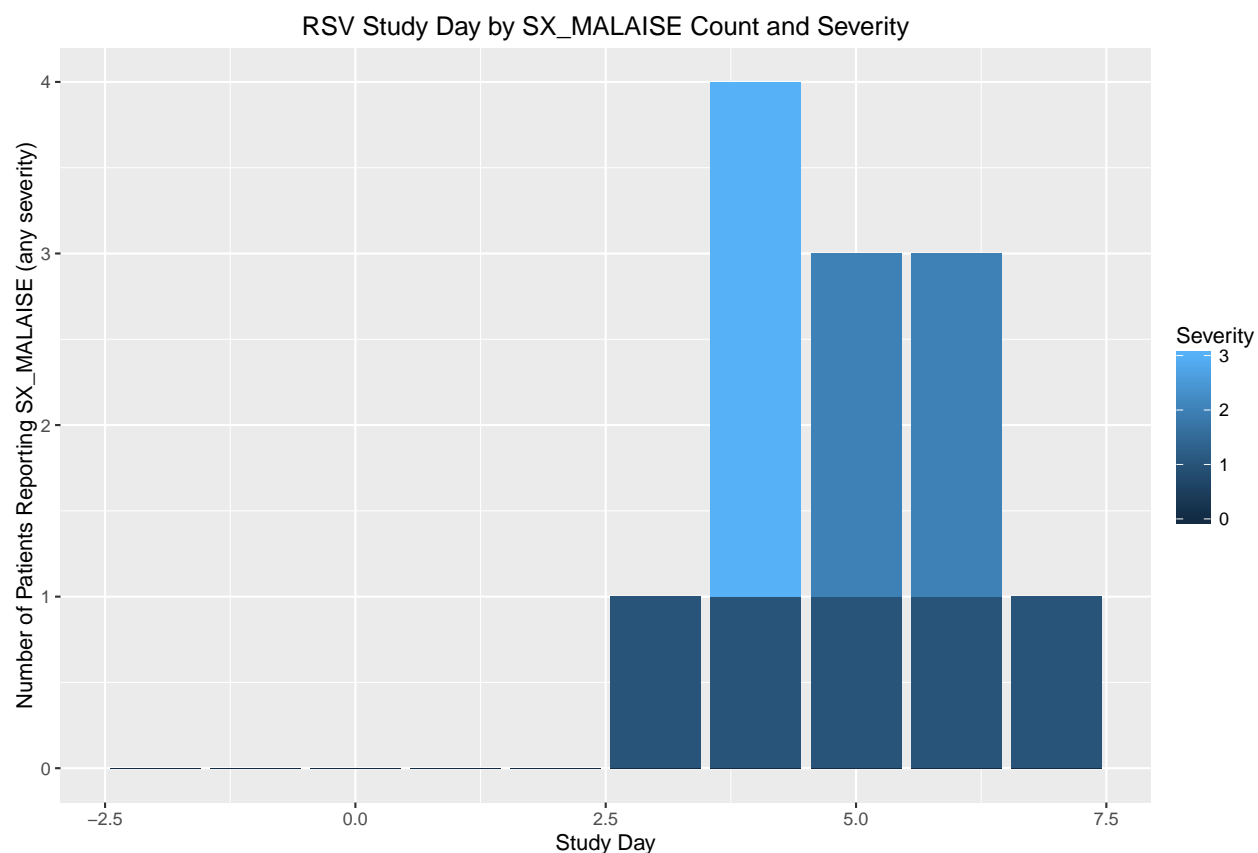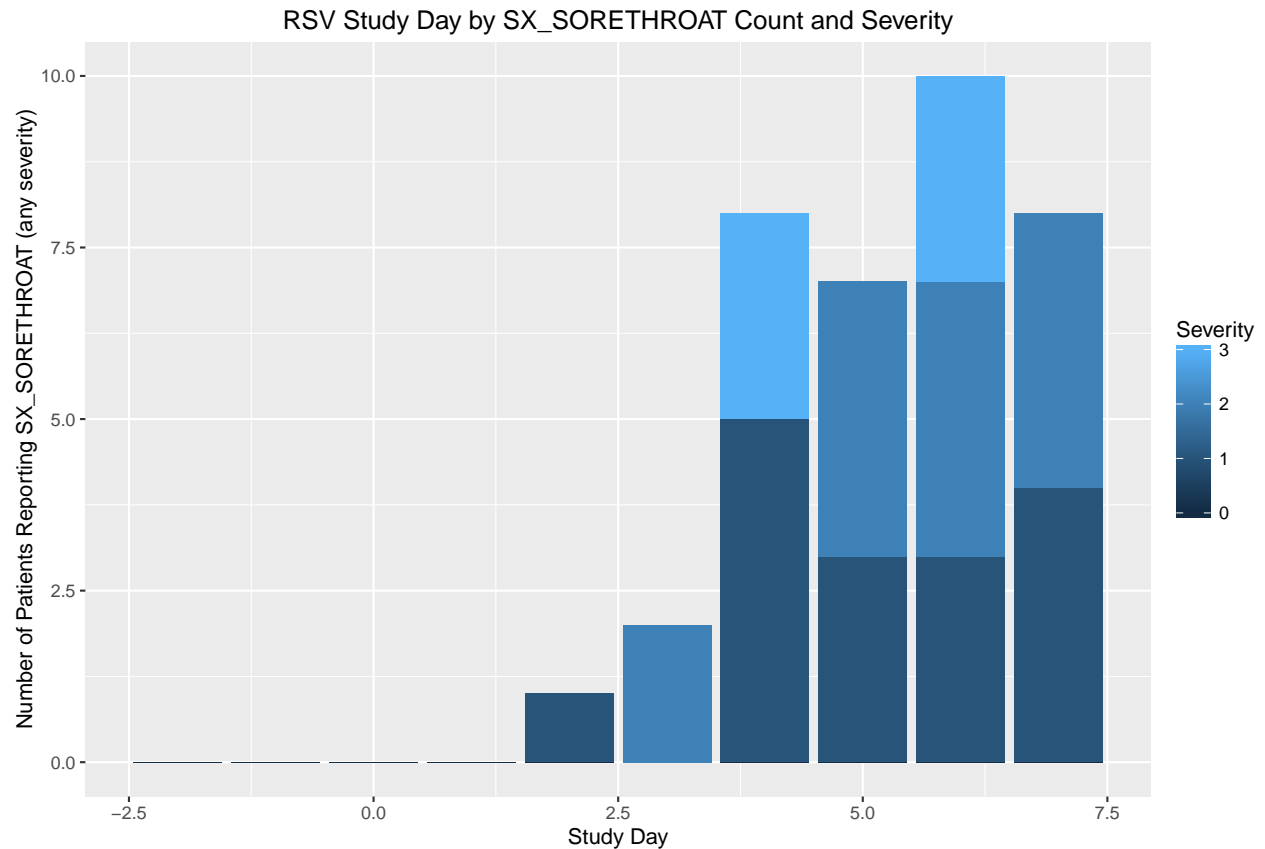
What phenotypes (symptoms/symptom combinations) are not seen pre-challenge?

> Rhinovirus: Sneeze
> H1N1: Cough
> RSV: Cough
> RSV: Malaise
> RSV: Myalgia
> RSV: Sore Throat

Look at removing patients who are asymptomatic for the entire study. Could we use them as shams?

```
z <- train_symptom_df %>%
  dplyr::rowwise() %>%
  dplyr::mutate(SUM_SCORE = sum(SX_HEADACHE,
                                SX_RUNNYNOSE,
                                SX_COUGH,
                                SX_MALAISE,
                                SX_MYALGIA,
                                SX_SNEEZE,
                                SX_SORETHROAT,
                                SX_STUFFYNOSE))
  z <- z %>% dplyr::ungroup() %>%
    dplyr::group_by(SUBJECTID,STUDYID) %>%
```

```
    dplyr::select(SUBJECTID,STUDYID,SUM_SCORE) %>%
  dplyr::summarize_each(dplyr::funs(sum)) %>%
  dplyr::filter(SUM_SCORE == 0)
  z %>% dim()
```

[1] 16 3

```
  z %>%
  pander(split.cells = 50, split.table = Inf)
```

| | | SUM_SCORE |
SUBJECTID	STUDYID	
13	Rhinovirus UVA	0
3014	DEE3 H1N1	0
3016	DEE3 H1N1	0
3022	DEE3 H1N1	0
3023	DEE3 H1N1	0
3024	DEE3 H1N1	0
4035	DEE4X H1N1	0
4038	DEE4X H1N1	0
5013	DEE5 H3N2	0
flu002	DEE2 H3N2	0
flu003	DEE2 H3N2	0
flu004	DEE2 H3N2	0
flu017	DEE2 H3N2	0
HRV10-013	Rhinovirus Duke	0
RSV004	DEE1 RSV	0
RSV019	DEE1 RSV	0

```
  z <- train_clinicl_df %>%
    dplyr::group_by(SUBJECTID,STUDYID) %>%
    dplyr::select(SUBJECTID,STUDYID,LOGSYMPTSCORE_SC3) %>%
  dplyr::summarize_each(dplyr::funs(sum)) %>%
  dplyr::filter(LOGSYMPTSCORE_SC3 == 0)
  z %>% dim()
```
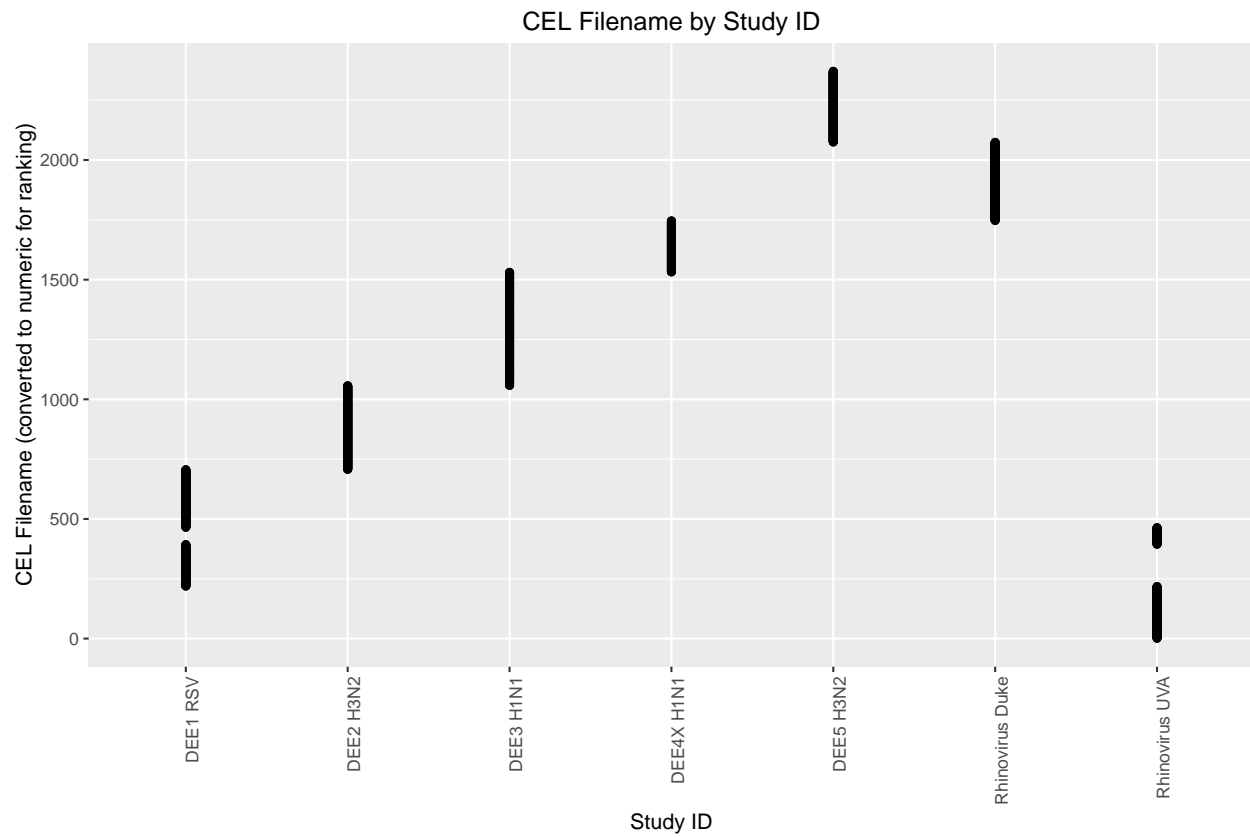
[1] 18 3

```
  z %>%
  pander(split.cells = 50, split.table = Inf)
```

| | | LOGSYMPTSCORE_SC3 |
SUBJECTID	STUDYID	
13	Rhinovirus UVA	0
3004	DEE3 H1N1	0
3014	DEE3 H1N1	0
3016	DEE3 H1N1	0
3022	DEE3 H1N1	0
3023	DEE3 H1N1	0

		LOGSYMPTSCORE_SC3
SUBJECTID	STUDYID	
3024	DEE3 H1N1	0
4035	DEE4X H1N1	0
4038	DEE4X H1N1	0
5013	DEE5 H3N2	0
flu002	DEE2 H3N2	0
flu003	DEE2 H3N2	0
flu004	DEE2 H3N2	0
flu017	DEE2 H3N2	0
HRV10-002	Rhinovirus Duke	0
HRV10-013	Rhinovirus Duke	0
RSV004	DEE1 RSV	0
RSV019	DEE1 RSV	0

Explore ID's of CEL files

CEL Filename by Study ID

Explore early TX