

Cutoff based on observed SBR:NMR ratio

Exclusion based on high quality data

- one row = one probability P,
- columns are minimum followed by sample percentiles 2.5%, 5%, 10%, 15%, 50%.

```
t(apply(percentile2,1,quantile,na.rm=TRUE,c(0,0.025,0.05,0.10,0.15,0.5)))
```

```
##           0%      2.5%      5%      10%      15%      50%
## 1%  0.2493324 0.4114587 0.4192604 0.4732101 0.4861460 0.6524495
## 2.5% 0.2768766 0.4287899 0.4328687 0.4900170 0.5059325 0.6773032
## 5%   0.3019553 0.4487959 0.4548431 0.5020789 0.5186885 0.6953818
## 7.5% 0.3175984 0.4564532 0.4693804 0.5131134 0.5294108 0.7072396
## 10%  0.3316621 0.4618311 0.4794762 0.5199703 0.5377465 0.7150664
## 15%  0.3544613 0.4720307 0.4968575 0.5332564 0.5516952 0.7296659
## 20%  0.3701759 0.4807520 0.5075180 0.5393984 0.5603432 0.7496284
```

```
ex.full.dat <- full_data %>% mutate(adj_sbr_unknown = ifelse(!is.na(adj_sbr_unknown),adj_sbr_unknown,SBR)
                                filter(!is.na(exclusion_notes))
```

```
ftb <- ex.full.dat$nLB + ex.full.dat$nSB
fsbr <- ex.full.dat$adj_sbr_unknown
flb <- ex.full.dat$nLB
fnmr <- ex.full.dat$NMR
```

```
#there are lots of missings in total birth
summary(ftb)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##         0   62619   143121  454648   427714  4189831      710
```

```
summary(flb)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##         0   60483   135820  433781   400133  4158445      624
```

```
summary(fsbr)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##    0.000   4.765    7.433   10.321   11.538   135.447      295
```

```
summary(fnmr)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##     0.00    5.40    9.80   12.15   15.70   86.84      200
```

```
fn <- length(ftb)
S <- 1000
fratio_s <- matrix(NA,ncol=S,nrow=fn)
for(i in 1:fn){
  sb_s <- rbinom(S,ftb[i],fsbr[i]/1000)
  nm_s <- rbinom(S,flb[i],fnmr[i]/1000)
  fratio_s[i,] <- sb_s/ftb[i]*1/(nm_s/flb[i])
}
fn # There are 2141 obs after first excluded.
```

```
## [1] 2141
```

```
fper_upper <- apply(fratio_s,1,quantile,na.rm = TRUE,c(0.95))  
sum(is.na(fper_upper))
```

```
## [1] 904
```

```
sum(fper_upper < 0.3,na.rm = T)
```

```
## [1] 75
```

```
sum(fper_upper < 0.5,na.rm = T)
```

```
## [1] 185
```

It seems that the approach works well. But the missing observation is a big problem. There are 904 missings among 2121 observations.