# Notes on calculating combined GAM estimates within the Rapid Assessment Method (RAM)

Mark Myatt

15 February 2020

# Contents

# 1 Background

A June 2019 article in Field Exchange advocates for the reporting of combined GAM estimate (i.e., a single prevalence estimate for GAM cases identified by WHZ and by MUAC). By January 2020, the SMART ENA software has included the calculation and reporting of this single GAM prevalence estimate.

This document reports on our current thining on how this single prevalence estimate can be calculated within the Rapid Assessment Method (RAM). Additionally, we explore how oedema cases can also be added in calculating this single GAM prevalence estimate.

# 2 Proposed approach

PROBIT gives a probability so we look to combining two probabilities:

$$P(GAM_{\text{MUAC}} \ \cup \ GAM_{\text{WHZ}}) \ = \ P(GAM_{\text{MUAC}}) \ + \ P(GAM_{\text{WHZ}})$$

However, the problem is that we do not have `independent` probabilities. We overestimate because the intersection gets counted twice. Therefore we need:

$$P(GAM_{\text{MUAC}} \cup GAM_{\text{WHZ}}) \ = \ P(GAM_{\text{MUAC}}) + P(GAM_{\text{WHZ}}) - P(GAM_{\text{MUAC}} \cap GAM_{\text{WHZ}})$$

We have the first two terms but not the third. We can estimate the third term from a 2 by 2 table:

| | $\texttt{WHZ} < -2$ | $\texttt{WHZ} \geq -2$ |
|---|---|---|
| $\texttt{MUAC} < 125$ | a | b |
| $\texttt{MUAC} \geq 125$ | c | d |

and

$$P(GAM_{\text{MUAC}} \ \cap \ GAM_{\text{WHZ}}) \ = \ \frac{a}{a \ + \ b \ + \ c \ + \ d}$$

We have a small sample size so the estimate will lack precision but I think that being "clever" and using something like:

$$P(GAM_{\text{MUAC}} \ \cap \ GAM_{\text{WHZ}}) \ = \ P(GAM_{\text{MUAC}}) \ \times \ P(GAM_{\text{WHZ}})$$

will not work as it assumes independence.

We can try to move forward with this hybrid method.

# 3 Worked example for combined GAM by WHZ and GAM by MUAC

We try the proposed approach in R using a dataset from Uganda.

```
## Read dataset from Uganda
x <- read.table(file = "data/ugan01.csv", header = TRUE, sep = ",")
```

```
## First 10 rows of the Uganda dataset
```

```
##     psu age sex weight height muac oedema   haz   waz   whz
## 1  9999  11   2    6.5   75.0   94      2  0.88 -2.47 -3.94
## 2  9999  18   1    5.5   67.5   94      2 -5.47 -5.51 -4.42
## 3  9999  12   2    5.2   63.3  100      2 -4.16 -4.40 -2.89
## 4  9999   8   1    6.3   69.8  100      2 -0.36 -2.83 -3.59
## 5  9999  17   1    6.2   68.5  101      2 -4.83 -4.64 -3.35
## 6  9999  27   1    8.0   78.0  101      2 -3.59 -3.92 -2.90
## 7  9999  24   2    6.2   68.4  103      2 -5.37 -4.92 -2.69
## 8  9999  19   1    7.0   73.0  103      2 -3.72 -4.02 -3.28
## 9  9999  12   2    6.0   67.4  104      2 -2.57 -3.36 -2.74
## 10 9999  11   1    5.6   66.2  105      2 -3.58 -4.46 -3.76
```

We then create case definitions as follows:

```
## Case definitions
x$gamWHZ <- ifelse(x$whz < -2, 1, 2)                   ## GAM by WHZ
x$gamMUAC <- ifelse(x$muac < 125, 1, 2)                ## GAM by MUAC
x$cGAM <- ifelse(x$whz < -2 | x$muac < 125, 1, 2)  ## GAM by WHZ and MUAC
```

```
## First 10 rows of the updated Uganda dataset contining case definitions
```

```
##    psu age sex weight height muac oedema   haz   waz   whz gamWHZ gamMUAC cGAM
## 1  9999  11   2    6.5   75.0   94      2  0.88 -2.47 -3.94      1       1    1
## 2  9999  18   1    5.5   67.5   94      2 -5.47 -5.51 -4.42      1       1    1
## 3  9999  12   2    5.2   63.3  100      2 -4.16 -4.40 -2.89      1       1    1
## 4  9999   8   1    6.3   69.8  100      2 -0.36 -2.83 -3.59      1       1    1
## 5  9999  17   1    6.2   68.5  101      2 -4.83 -4.64 -3.35      1       1    1
## 6  9999  27   1    8.0   78.0  101      2 -3.59 -3.92 -2.90      1       1    1
## 7  9999  24   2    6.2   68.4  103      2 -5.37 -4.92 -2.69      1       1    1
## 8  9999  19   1    7.0   73.0  103      2 -3.72 -4.02 -3.28      1       1    1
## 9  9999  12   2    6.0   67.4  104      2 -2.57 -3.36 -2.74      1       1    1
```

```
## 10 9999  11   1    5.6   66.2  105      2 -3.58 -4.46 -3.76      1       1    1
```

## 3.1 Combined GAM estimation using the classical approach

Calculating for prevalence using the classical approach we get:

```
## Classic prevalence for GAM by MUAC
round(prop.table(table(x$gamMUAC))[1] * 100, 2)
```

```
##    1
## 13.8
```

```
## Classic prevalence for GAM by WHZ
round(prop.table(table(x$gamWHZ))[1] * 100, 2)
```

```
##    1
## 9.05
```

```
## Classic prevalence for GAM by WHZ and MUAC
round(prop.table(table(x$cGAM))[1] * 100, 2)
```

```
##    1
## 15.5
```

We can test whether GAM cases by MUAC and GAM cases by WHZ are independent.

```
## Test if the two case definitions are independent
chisq.test(table(x$gamMUAC, x$gamWHZ))
```

The chi-square test has a *p-value* of $8.8108361 \times 10^{-74}$ indicating that the two case definitions are not independent.

## 3.2 Combined GAM estimation using proposed hybrid approach

We then proceed with our proposed hybrid approach using simple PROBIT prevalence estimation.

```
## Simple PROBIT prevalence for GAM by MUAC
pMUAC <- pnorm(125, mean(x$muac), sd(x$muac))
```

```
## [1] 13.65
```

```
## Simple PROBIT prevalence for GAM by WHZ
pWHZ <- pnorm(-2, mean(x$whz), sd(x$whz))
```

```
## [1] 7.61
```

```
## Estimate the UNION probability
pUNION <- table(x$gamMUAC, x$gamWHZ)[1,1] / sum(table(x$gamMUAC, x$gamWHZ))
```

```
## [1] 7.35
```

```
## Estimate of GAM by MUAC and WHZ by PROBIT
round((pMUAC + pWHZ - pUNION) * 100, 2)
```

```
## [1] 13.91
```

# 4 Extending the approach to include oedema cases

Using the same concepts, we can extend the approach to include oedema cases. For this we need to take into account not only the intersection between GAM by MUAC and GAM by WHZ but also the intersection of GAM by MUAC and oedema cases, the intersection of GAM by WHZ and oedema cases, and the intersection between GAM by MUAC, GAM by WHZ and oedema cases as follows:

$$\begin{aligned}
P(GAM_{\text{MUAC}} \cup GAM_{\text{WHZ}} \cup SAM_{\text{oedema}}) = \\
& P(GAM_{\text{MUAC}}) + P(GAM_{\text{WHZ}}) + P(SAM_{\text{oedema}}) \\
& - [P(GAM_{\text{MUAC}} \cap GAM_{\text{WHZ}}) - P(GAM_{\text{MUAC}} \cap GAM_{\text{WHZ}} \cap SAM_{\text{oedema}})] \\
& - [P(GAM_{\text{MUAC}} \cap SAM_{\text{oedema}}) - P(GAM_{\text{MUAC}} \cap GAM_{\text{WHZ}} \cap SAM_{\text{oedema}})] \\
& - [P(GAM_{\text{WHZ}} \cap SAM_{\text{oedema}}) - P(GAM_{\text{MUAC}} \cap GAM_{\text{WHZ}} \cap SAM_{\text{oedema}})] \\
& - P(GAM_{\text{MUAC}} \cap GAM_{\text{WHZ}} \cap SAM_{\text{oedema}})
\end{aligned}$$

Simplifying this we get:

$$\begin{aligned}
P(GAM_{\text{MUAC}} \cup GAM_{\text{WHZ}} \cup SAM_{\text{oedema}}) = \\
& P(GAM_{\text{MUAC}}) + P(GAM_{\text{WHZ}}) + P(SAM_{\text{oedema}}) \\
& - P(GAM_{\text{MUAC}} \cap GAM_{\text{WHZ}}) \\
& - P(GAM_{\text{MUAC}} \cap SAM_{\text{oedema}} \\
& - P(GAM_{\text{WHZ}} \cap SAM_{\text{oedema}}) \\
& + (2 \times P(GAM_{\text{MUAC}} \cap GAM_{\text{WHZ}} \cap SAM_{\text{oedema}}))
\end{aligned}$$

# 5 Worked example for combined GAM by WHZ, GAM by MUAC and oedema

We try the proposed approach in R using the same dataset from Uganda used in the previous example. We already have an oedema variable that is coded into 1 for an oedema case and 2 for a non-oedema case. We just need to re-define the `cGAM` variable to include oedema cases.

```r
## Case definition for combined GAM including oedema cases
x$cGAM <- ifelse(x$whz < -2 | x$muac < 125 | x$oedema == 1, 1, 2)

## First 10 rows of the updated Uganda dataset contining case definitions
```

```
##     psu age sex weight height muac oedema   haz   waz   whz gamWHZ gamMUAC cGAM
## 1  9999  11   2    6.5   75.0   94      2  0.88 -2.47 -3.94      1       1    1
## 2  9999  18   1    5.5   67.5   94      2 -5.47 -5.51 -4.42      1       1    1
## 3  9999  12   2    5.2   63.3  100      2 -4.16 -4.40 -2.89      1       1    1
## 4  9999   8   1    6.3   69.8  100      2 -0.36 -2.83 -3.59      1       1    1
## 5  9999  17   1    6.2   68.5  101      2 -4.83 -4.64 -3.35      1       1    1
## 6  9999  27   1    8.0   78.0  101      2 -3.59 -3.92 -2.90      1       1    1
## 7  9999  24   2    6.2   68.4  103      2 -5.37 -4.92 -2.69      1       1    1
## 8  9999  19   1    7.0   73.0  103      2 -3.72 -4.02 -3.28      1       1    1
## 9  9999  12   2    6.0   67.4  104      2 -2.57 -3.36 -2.74      1       1    1
## 10 9999  11   1    5.6   66.2  105      2 -3.58 -4.46 -3.76      1       1    1
```

## 5.1 Combined GAM estimation using the classical approach

Calculating for oedema and combine GAM prevalence using the classical estimator we get:

```r
## Classic prevalence for oedema cases
round(prop.table(table(x$oedema))[1] * 100, 2)
```

```
##    1
## 0.23
```

```r
## Classic prevalence for GAM by WHZ and MUAC and oedema cases
round(prop.table(table(x$cGAM))[1] * 100, 2)
```

```
##     1
## 15.61
```

## 5.2 Combined GAM estimation using proposed hybrid approach

We then proceed with our proposed hybrid approach using simple PROBIT prevalence estimation to include oedema cases.

```r
## Estimate the probability of oedema cases
pOedema <- prop.table(table(x$oedema))[1]

## Estimate union probability for WHZ and oedema
pOedemaWHZ <- table(x$gamWHZ, x$oedema)[1,1] / sum(table(x$gamWHZ, x$oedema))
```

```
## [1] 0.11
```

```r
## Estimate union probability for MUAC and oedema
pOedemaMUAC <- table(x$gamMUAC, x$oedema)[1,1] / sum(table(x$gamMUAC, x$oedema))
```

```
## [1] 0.11
```

```r
## Create variable for GAM meeting WHZ, MUAC and oedema criteria
x$gamUnion <- ifelse(x$oedema == 1 & x$gamWHZ == 1 & x$gamMUAC == 1, 1, 2)

## Estimate union probability for MUAC, WHZ and oedema
pGAMunion <- prop.table(table(x$gamUnion))[1]
```

```
##    1
## 0.11
```

```r
## Estimate combined GAM including oedema cases
round((pMUAC + pWHZ + pOedema - pUNION - pOedemaWHZ -
         pOedemaMUAC + (2 * pGAMunion)) * 100, 2)
```

```
##     1
## 14.14
```