Rapid Assessment Method

Development Update : Number 1

Introduction

This document reports on RAM development activities. This update reports desk-based work on:

- Improvements to the performance of the PROBIT estimator for GAM.
- Extension of the PROBIT indicator to include SAM in order to provide a needs assessment estimate for CMAM programs. This is an extension to RAM as originally proposed.

PROBIT for GAM : Introduction

The PROBIT estimator takes two parameters:

> **Location (central tendency) :** The location could be the mean or the median or some other estimator of the centre of the distribution of MUAC values.
>
> **Dispersion (or spread) :** The dispersion could be the standard deviation (SD) or some other measure of the spread of the distribution of MUAC values about centre of the distribution of MUAC values.

The mean and the SD are the most common measures of location and dispersion. They have the advantage of being easy to calculate because they do not require data to be sorted. The have the disadvantage of lacking *robustness* to outliers (i.e. extreme high or low values). This problem effects the SD more than the mean because its calculation involves squaring deviations from the mean and this gives more weight to extreme values. Lack of robustness is most problematic with small sample sizes as we have with RAM.

There are two ways around this problem :

1. Transform the data to bring outliers closer to the centre of the distribution. This only works well when outliers occur to one side or the centre of the distribution.
2. Use *robust estimators* of location and dispersion. These are estimators for location and dispersion that limit the effect of outliers.

This document reports upon tests of the PROBIT estimator using transformations and robust estimators of location and dispersion for estimating GAM prevalence.

PROBIT for GAM : Testing

The robust measures of location tested were:

Median : The value separating the higher half of a sample (or population or probability distribution) from the lower half of the sample. The median of a list of numbers can be found by arranging all the observations from lowest value to highest value and picking the middle one. If there is an even number of observations then the median is the mean of the two middle values.

Midhinge (TTM) : The mean of the first and third quartiles. In exploratory data analysis (EDA) methods the midhinge is complemented by the interquartile range (IQR, see below) as the measure of dispersion.

Tukey Trimean : This is a weighted average of the median and the two quartiles:

$$TTM = \frac{Q1 + 2 \times MEDIAN + Q3}{4}$$

The TTM combines the median's emphasis on centre values with the midhinge's attention to the extreme values.

The robust estimators of dispersion tested were:

MAD : The median absolute deviation (MAD) is a robust measure of dispersion that is calculated as the median of the absolute values of the deviation of each observation from the sample median. The MAD can be used to (robustly) estimate the standard deviation (SD) for a normally distributed variable:

$$SD = 1.48260 \times MAD$$

IQR : The interquartile range is the distance between the upper and lower quartiles. This is the range of the middle 50% of the values. The IQR can be used to (robustly) estimate the standard deviation (SD) for a normally distributed variable using:

$$SD = \frac{IQR}{1.34898}$$

Each candidate estimator was tested using computer-based simulation (as outlined in the RAM proposal document) with a sample size of $n = 192$ with 224,000 simulated surveys. The classic approach was also tested with $n = 544$ (this is the largest sample size mentioned in the SMART manual) with 224,000 simulated surveys. Results of testing are presented in the table below:

PERFORMANCE OF RAM/PROBIT CANDIDATE ESTIMATORS FOR GAM PREVALENCE

Method	Location	Dispersion	Error (%)	Rel. Prec. (%)
PROBIT	Mean	SD	0.8667	23.99
	Mean (transformed)	SD (transformed)	0.7321	24.05
	Median	MAD * 1.42860	0.1852	24.58
	Median	IQR / 1.34898	0.0670	24.66
	Tukey's Trimean	IQR / 1.34898	0.1059	24.62
	Mid-hinge	IQR / 1.34898	0.1947	24.58
CLASSIC	NA	NA	-0.0006	27.22

The results are presented as:

Error : This is the mean error associated with an estimator. If (e.g.) true prevalence is 10% and in ten surveys we found the following estimates of prevalence:

```
{10.35 10.23 9.43 10.90 10.77 11.08 9.65 10.06 11.25 11.04}
```

Then the individual errors (true prevalence – estimated prevalence) are:

```
{-0.35 -0.23 0.57 -0.90 -0.77 -1.08 +0.35 -0.06 -1.25 -1.04}
```

The mean error is the average of these values (i.e. -0.476 in the example data). This is a measure of the bias of an estimator. An unbiased estimator will have a mean error of zero. It is possible to use mean error to adjust estimates:

$$\text{corrected estimate} = \text{estimate} + \text{mean error}$$

Applying this to the example data gives:

```
{9.88 9.75 8.95 10.42 10.29 10.60 9.17 9.58 10.77 10.10.56}
```

The individual errors (true prevalence – estimated prevalence) are:

```
{0.12 0.25 1.05 -0.42 -0.29 -0.60 0.83 -0.42 -0.77 -0.56}
```

The mean error is now zero. We are looking to minimise error or to remove the need to use such a correction.

Rel. Prec. : This is relative precision. It is calculated as:

$$\text{Relative precision} = \frac{1}{\text{estimate}} \times \frac{(UCL - \text{estimate}) + (\text{estimate} - LCL)}{2}$$

This is a measure of precision that is expressed as a proportion of the estimate. If (e.g.) a survey returned the following result:

prevalence = 10.13% (95% CI = 7.30 - 13.58)

The relative precision would be:

$$\text{Relative precision} = \frac{1}{10.13} \times \frac{(13.58 - 10.13) + (10.13 - 7.30)}{2}$$

$$\text{Relative precision} = 0.31 = 31\%$$

In the table of results we present the mean relative precision for all simulated surveys as a summary of performance. Smaller values are better.

The PROBIT estimator using the median and IQR has a very small error and a relative precision that is only a little worse than the other PROBIT estimators. This is probably the best estimator to use for estimation of GAM. This is an approximately unbiased estimator for GAM with precision better than a SMART survey with a much larger sample size.

Figure U1.1 shows error by true prevalence at four sample sizes for the PROBIT estimator using the median and IQR. *Figure U2.2* shows relative precision at four sample sizes for the PROBIT estimator using the median and IQR. These results indicate that RAM surveys should aim for an effective sample size of about n = 192.

Figure U1.1 : Error by true prevalence at the tested sample sizes (GAM)

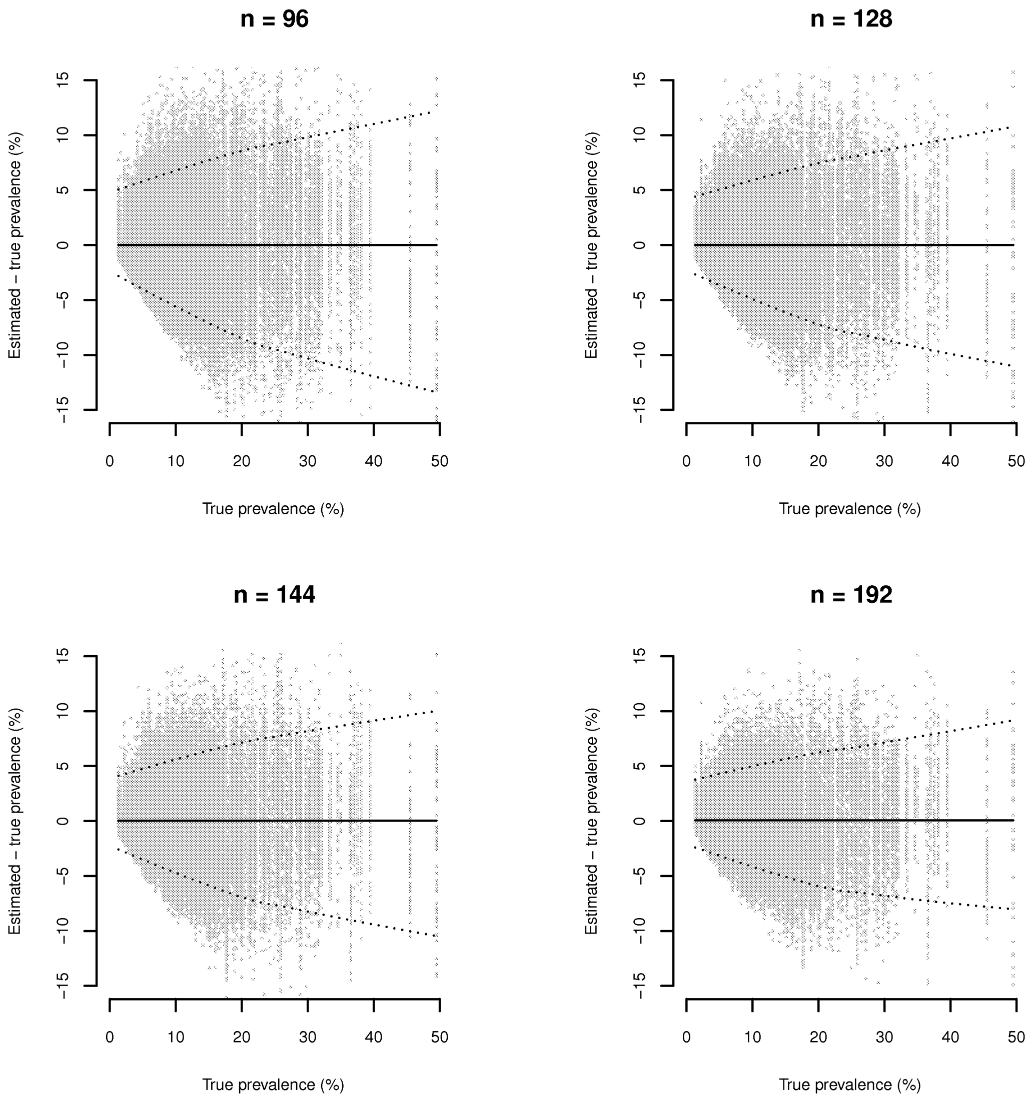
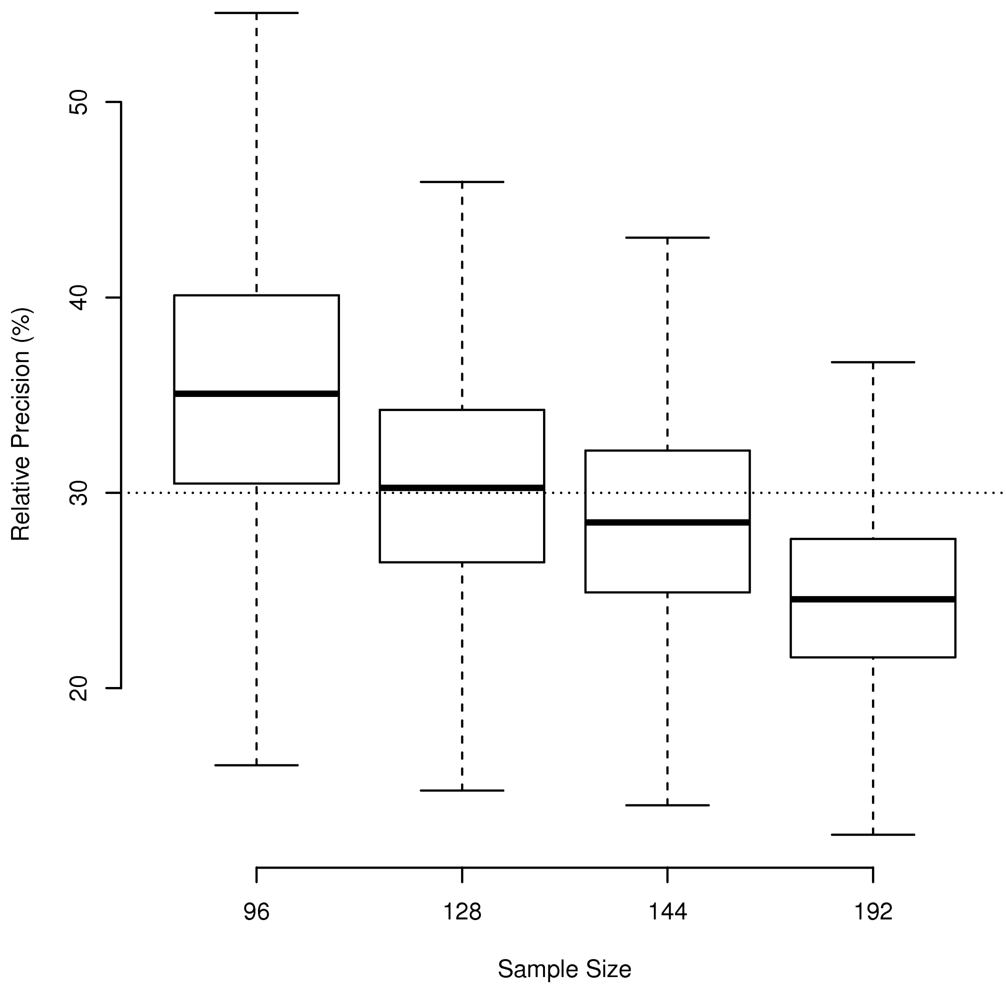


Figure U1.2 : Relative precision at the tested sample sizes (GAM)



PROBIT for SAM

A similar testing procedure was used for estimating the prevalence of SAM for estimating need for CMAM services. This is an extension to RAM as originally proposed.

Results of the testing are:

PERFORMANCE OF RAM/PROBIT CANDIDATE ESTIMATORS FOR SAM PREVALENCE

Method	Location	Dispersion	Error (%)	Rel. Prec. (%)
PROBIT	Mean	SD	0.1225	33.62
	Mean (transformed)	SD (transformed)	0.0438	33.47
	Median	MAD	0.4484	34.61
	Median	IQR / 1.34898	0.3735	34.72
	Tukey's Trimean	IQR / 1.34898	0.4366	34.68
	Mid-hinge	IQR / 1.34898	0.4079	34.65
CLASSIC	NA	NA	0.0022	65.03

The PROBIT estimator using the data transformed towards normality has the lowest error. This estimator could be used. It is also possible to use the PROBIT estimator using the median and IQR using a correction factor for estimating need:

$$need = (estimate + 0.3735) \times population_{6-59 Months} \times 1.6 \times expected\ coverage$$

since this is always a very approximate calculation.

This estimator has better precision than a SMART survey with a much larger sample size.

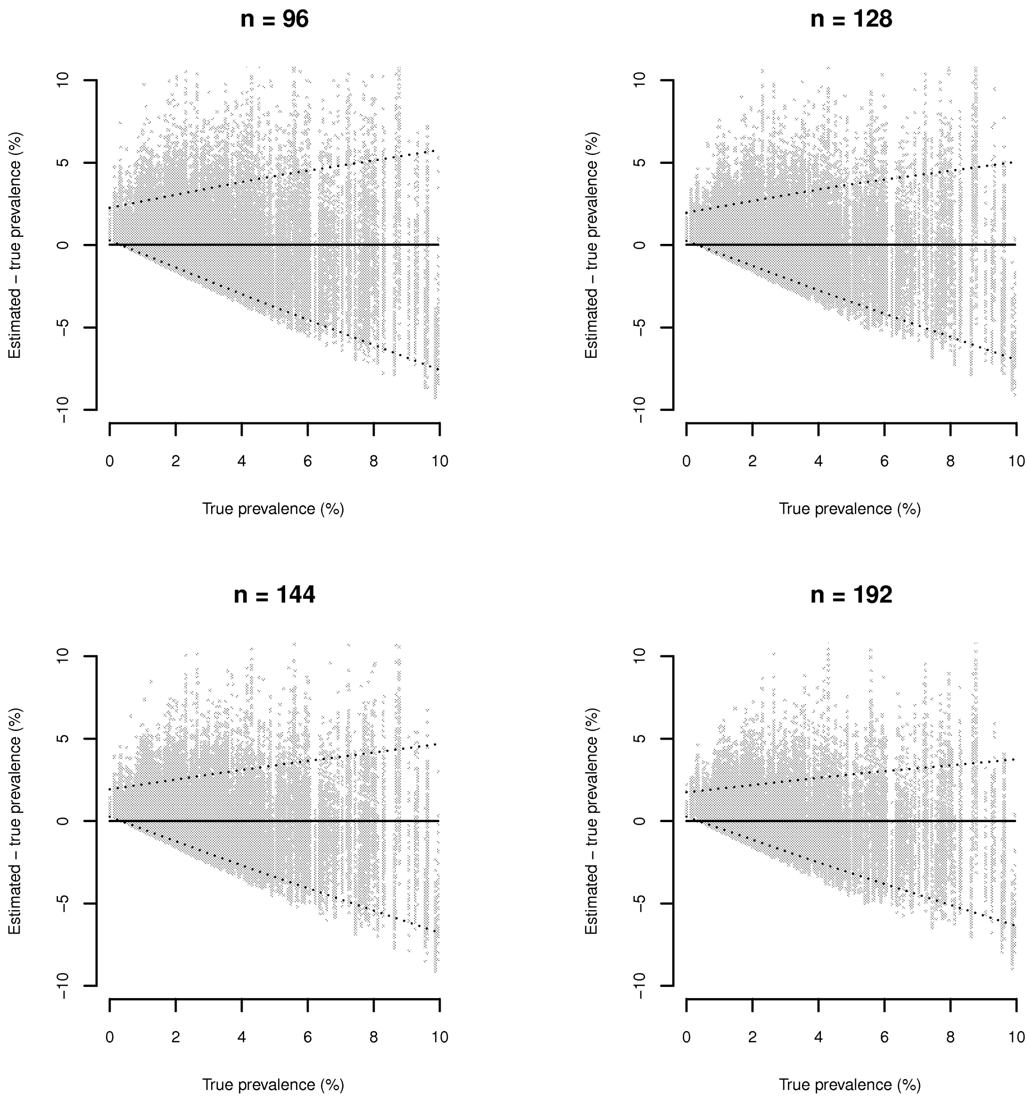
Figure U1.3 shows error by true prevalence at four sample sizes for the PROBIT estimator using the median and IQR. *Figure U2.4* shows relative precision at four sample sizes for the PROBIT estimator using the median and IQR. These results also indicate that RAM surveys should aim for an effective sample size of about n = 192.

Note that:

MAM = GAM – SAM

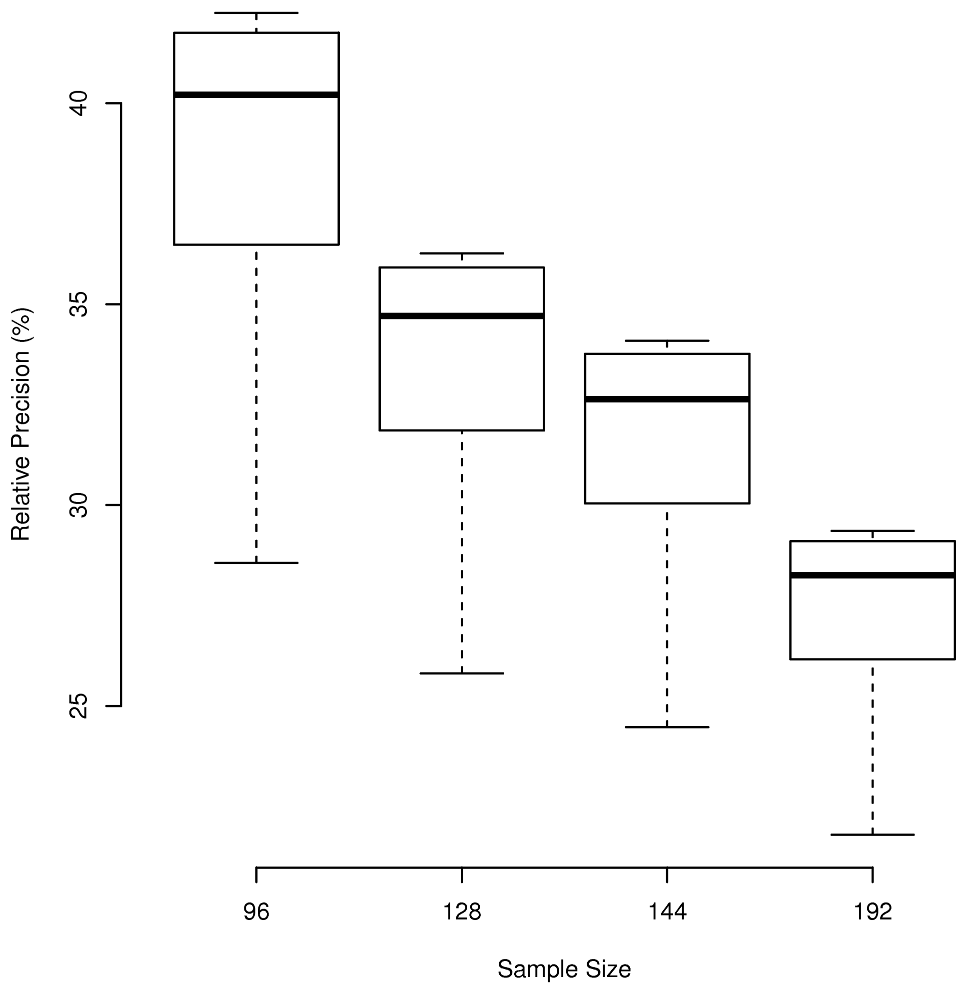
This means that the proposed method can provide estimates of GAM, MAM, and SAM.

Figure U1.3 : Error by true prevalence at the tested sample sizes (SAM)



Results shown for SAM prevalence < 10% only

Figure U1.4 : Relative precision at the tested sample sizes (SAM)



Results shown for SAM prevalence < 10% only

Summary

The RAM development work reported here supports the following conclusions:

- RAM surveys using the PROBIT estimator using the median and IQR can provide accurate and usefully precise estimates of GAM prevalence.

- RAM surveys using the PROBIT estimator using the median and IQR can provide estimates of SAM prevalence that may be useful for estimating need for CMAM services.

- RAM surveys should aim for an effective sample size of n = 192.

It is suggested that upcoming RAM pilot surveys use the PROBIT estimator using the median and IQR with a sample size of n = 192.

Why choose *n* = 192?

The sample size n = 192 provided useful precision at a reasonable small sample size.

Using n = 192 simplifies partition of the sample between sampling units because it has a lot of whole number divisors (i.e. 1, 2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 64, 96, 192) and the list of its whole number divisors includes four (this is important as the proposed sampling method uses the “QTR” community segmentation procedure). A sample of size n = 192 could (e.g.) be collected as 4 children from each of 4 “QTR” segments of 12 communities.