Is there a systematic bias in estimates of programme coverage returned by SQUEAC coverage assessments?

By Mark Myatt and Ernest Gueverra

Mark Myatt is a consultant epidemiologist. His areas of expertise include infectious disease, nutrition and survey design. He worked with FANTA, Valid International, the US Centers for Disease Control and Prevention, Tufts University, Action Against Hunger UK, Concern Worldwide and the Coverage Monitoring Network as the lead developer of the SQUEAC coverage assessment method.

Ernest Guevarra leads Valid International's coverage assessment and surveys team. He has formal training as a physician and a public health practitioner. Most recently he has worked in Sierra Leone, Niger, Sudan, Ethiopia and Ghana. He is the lead developer of SQUEAC, SLEAC and S3M coverage assessment methods at Valid International.

Location: *Global*

What we know: Used since 2012, the Semi-Quantitative Evaluation of Access and Coverage (SQUEAC) coverage assessment method employs both qualitative and quantitative methods to identify key barriers to access and estimate coverage of therapeutic feeding programmes (TFP) and, to a lesser extent, supplementary feeding programmes (SFP).

What this article adds: A recent article by Isanaka et al (2018) on SQUEAC implementation in Niger suggests that the analysis required is technically demanding and in part relies on subjective estimates of programme coverage. With typical operational capacities this will cause SQUEAC assessments to systematically overestimate coverage. This article investigates the risk of systematic bias by analysing a database of 304 SQUEAC coverage assessment reports and data from 29 countries (2009-2017). There is a tendency for the 'prior' (an informed guess about programme coverage) to overestimate coverage when the true coverage is low and underestimate coverage when the true coverage is high. There is an equal risk of the prior overestimating and underestimating coverage (i.e. no systematic bias). Problems were detected in 7.3% of the SQUEAC assessments reviewed but this led to coverage estimates with poor precision in only 2.25% of assessments. The use of untrained staff and failure to use SQUEAC processes, methods and tools correctly is likely to increase this risk. The authors conclude there is no evidence of general and systematic overestimation of coverage using SQUEAC and that the risk of the method yielding estimates with poor precision is low. A key lesson from the Isanaka et al (2018) SQUEAC experience is the importance of using both properly trained staff and using SQUEAC processes, methods and tools correctly.

Background

A recent article identifies a potentially serious problem with coverage estimates made using the SQUEAC coverage assessment method. The article:

Isanaka S, Hedt-Gauthier BL, Grais RF, Allen BG, *Estimating program coverage in the treatment of severe acute malnutrition: a comparative analysis of the validity and operational feasibility of two methods*, Population Health Metrics, 2018,16:100,1-9

A summary of the article is given in Box 1. The full version of the article is available from: **https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-018-0167-3**

Coverage estimates made by SQUEAC rely on condensing data collected from a variety of sources using a range of methods to make an informed guess about the level of coverage that a program is achieving. This informed guess is known as the *prior*. The prior is used to inform the design of a small sample coverage survey. The prior is also combined with the coverage survey data, known as the *likelihood*, to provide an estimate of the coverage that a program is achieving using a widely accepted statistical technique known as *conjugate analysis*. A problem with this approach is that a very poorly specified prior can result in a biased estimate of coverage. A prior that is much higher than the true coverage can lead to an upwardly biased estimate of coverage. A prior that is much lower than the true coverage can lead to a downwardly biased estimate of coverage. These situations are known as prior-likelihood conflicts. If a *prior-likelihood conflict* is detected, the results of the conjugate analysis are discarded and a coverage estimate is made using the survey data alone. This estimate will not be biased but may lack precision (i.e. have a wide 95% credible interval) due to the small sample size used in the coverage survey. The article by Isanaka et al. (2018) identifies prior-likelihood conflicts as a weakness of the SQUEAC coverage assessment method which leads the method to systematically overestimate coverage.

The issue of prior-likelihood conflicts is not new. It is covered at some length in the SQUEAC technical reference. A formal test for prior-likelihood conflicts has been provided by the **BayesSQUEAC** calculator for several years. The specific case of untrained staff producing an unrealistically optimistic and overly strong prior, as reported in the article be Isanaka et al. (2018), is presented as a case-study in the SQUEAC technical reference. The article by Isanaka et al. (2018) confirms the existence of a problem that is frankly admitted, discussed, and cautioned against in SQUEAC documentation and training. This should not, however, prevent us from taking this criticism of the SQUEAC method very seriously. It is possible that there is a serious problem with the SQUEAC method which is leading to a general and systematic failure to identify programs failing to meet coverage standards and leaving many vulnerable children untreated. This issue is investigated in this article.

Method

A database was created from SQUEAC coverage assessment reports and SQUEAC coverage assessment data provided by the Coverage Monitoring Network and VALID International. Reports and data for n = 304 SQUEAC coverage assessments from 29 countries undertaken between 2009 and 2017 were available. Only data from SQUEAC coverage assessments which completed a SQUEAC stage III coverage estimation survey (n = 274) are included in the analysis reported here.

For each SQUEAC coverage assessment, the mode of the prior was calculated as:

$$prior\ mode = \frac{\alpha_{Prior}\text{-}1}{\alpha_{Prior+\beta prior}\text{-}2}$$

The numerator ($N_{Likelihood}$) and denominator ($D_{Likelihood}$) for the likelihood mode were calculated for the principal coverage estimator (i.e. point, period, or single coverage) reported in the SQUEAC coverage assessment report:

$$likelihood\ mode = \frac{N_{Likelihood}}{D_{Likelihood}}$$

The relationship between the prior modes and the likelihood modes was explored by calculating, plotting and summarising the difference:

$$prior\ mode\ \text{-}\ likelihood\ mode$$

and by plotting the prior modes against the likelihood modes.

The strength of the linear association between the prior modes and the likelihood modes was assessed using the Pearson correlation coefficient. Ordinary least squares linear regression was used to determine the slope of the line that best described the relationship between the prior modes and the likelihood modes.

For each SQUEAC coverage assessment, prior-likelihood conflicts were detected using a testing approach. Two-by-two tables were constructed with cells:

$\lVert\alpha_{prior}\text{-}1\rVert$	$\lVert\beta_{prior}\text{-}1\rVert$
$N_{Likelihood}$	$D_{Likelihood}\text{-}N_{Likelihood}$

and Fisher's exact test of independence calculated for each of the constructed tables. The null hypothesis for Fisher's exact test is:

H_0 : prior mode = likelihood mode

A prior likelihood conflict is detected when this null hypothesis is rejected. This occurs when either of the alternative hypotheses:

H_A : prior mode >> likelihood mode

H_A : prior mode << likelihood mode

is more consistent with the observed data than the null hypothesis. That is, a two-sided test was used. The ">>" and "<<" are used to indicate prior modes that are greater than and less than would be expected by chance alone. This is a similar procedure to applying the z-test used by Isanaka et al. (2018) and in the **BayesSQUEAC** calculator that is used in most SQUEAC coverage assessments. Fisher's exact test was used to avoid issues with small sample sizes and very unequal distribution of data within tables giving rise to small expected numbers which would be problematic if approximate methods such as the z-test and the chi-squared test were used. A two-sided p-value of $p < 0.05$ was taken as evidence of a prior-likelihood conflict.

The half-width of 95% confidence intervals (E) for the likelihood modes (p) were calculated using the normal approximation and applying a finite population correction (i.e. because severe acute malnutrition is a rare condition) for each SQUEAC coverage assessment for which a prior-likelihood conflict was detected:

$$E = 1.96 \times \sqrt{\left(\frac{p\ (1\text{-}p)}{n}\right)} \times \sqrt{\left(\frac{Population\ Size\text{-}D_{Likelihood}}{Population\ Size\text{-}1}\right)}$$

The population size used to calculate the finite population correction (i.e. 600) was calculated assuming an overall population of 100,000 persons with 20% aged between 6 and 59 months and a 3% prevalence of severe acute malnutrition (SAM). These are conservative assumptions. It was not necessary to use the prevalence of moderate acute malnutrition (MAM) to calculate the finite population correction as no prior-likelihood conflicts were found in assessments of supplementary feeding programmes (SFP).

The *relative precision* achieved using the likelihood data alone was calculated as:

$$relative\ precsion = \frac{E}{p}$$

A coverage estimate of 50% (p) with a 95% confidence interval of ± 10% (E) has relative precision:

$$relative\ precsion = \frac{E}{p} = \frac{10\%}{50\%} = 20\%$$

The resulting relative precision was compared to the relative precision that would have been achieved by a standard Expanded Program of Immunisation (EPI) '30 x 7' coverage survey with the same point estimate of the coverage proportion (p), a sample size of n = 210 and a survey design effect of 2.0. This relative precision was used as the 'gold standard' for the precision of methods assessing the coverage of child survival programmes. A SQUEAC assessment was classified as 'failing' if a prior-likelihood conflict was detected and the relative precision of the coverage estimate made using the likelihood data alone was worse than this gold standard.

Results

Table 1 presents a description of the study database. Figure 1 shows the distribution of the differences between the prior modes and likelihood modes. The median difference was −0.97% (IQR = −8.31%; +8.39%). The differences were normally distributed (Shapiro-Wilk Normality Test p = 0.6287) about a central value close to zero (mean = −0.67%, 95% CI = −2.25%; +0.90%).

Box 1 Summary of the article by Isanaka et al. (2018

The article by Isanaka et al. (2018) published in Population Health Metrics recognises that assessing the coverage of nutrition programs is challenging due to the low prevalence of disease and selective entry criteria. It also recognises that SQUEAC is a "step forward in coverage assessment of therapeutic feeding programs" and can "simultaneously identify barriers to accessing care and estimate program coverage". It notes, however, that "the validity of certain methodological elements has been the subject of debate". The methodological elements in question revolve about the use of a Bayesian conjugate analysis to improve the precision of coverage estimates made using small sample sizes. The concern is not that Bayesian approaches are generally invalid but that the approach is beyond the technical capacity of staff employed by NGOs, UNOs, and ministries of health and that its use in the wrong hands will lead to (worst case) systematic overestimation of coverage and (best case) coverage estimates with very poor precision.

Isanaka et al. (2018) investigate this issue by comparing the results of a SQUEAC coverage assessment performed by untrained persons against a two-stage cluster sample survey with a spatially stratified first stage selecting communities and active and adaptive case-finding in the second stage. The comparison method employed is very similar to the method used by SQUEAC stage III likelihood surveys. The only difference being that a larger sample size is used. This means that any substantial differences found between the SQUEAC results and the survey results will be due to the untrained staff doing a poor job of specifying the prior used in the Bayesian conjugate analysis.

The article reports that priors produced by untrained staff and by untrained community members led to upwardly biased coverage estimates. Point estimates of coverage made from the likelihood survey data alone were similar to those made from the larger two-stage cluster survey. This means that the problem is with the Bayesian prior produced by untrained staff and untrained community members being too optimistic and too strong (i.e. overly certain). The reported biases were, however, consistently detected using standard SQUEAC diagnostic methods and SQUEAC software (i.e. plots and tests in BayesSQUEAC) for detecting prior-likelihood conflicts. Coverage estimates made using a prior produced by trained staff was in agreement with that made by the two-stage cluster sample.

The authors conclude that SQUEAC is technically demanding and should only be used when the appropriate technical capacity is available. They also question the validity of the methods used by SQUEAC to produce priors when they are used in capacity limited settings.

Table 1 The study database

Item	Description	n
Number of records	Total number of records in the study database	304
Excluded	No stage III : Not required:	17
	No stage III : Suspected patchy coverage	5
	No stage III : Poor security / access	6
	No stage III : Very low SAM prevalence	2
Included	Number of records included in the analysis	274
Coverage type*	Point coverage	199
	Period coverage	70
	Single coverage	5
Assessed programme**	OTP	255
	SFP	19

* Point coverage measures case-finding and recruitment; Period coverage measures case-finding, recruitment and retention but overestimates coverage; Single coverage adjusts period coverage removing bias by including an estimate of the number of recovering cases in the community in the denominator.

** OTP = Outpatient Therapeutic Program treating cases of severe acute malnutrition; SFP = Supplementary Feeding Program treating cases of moderate acute malnutrition.

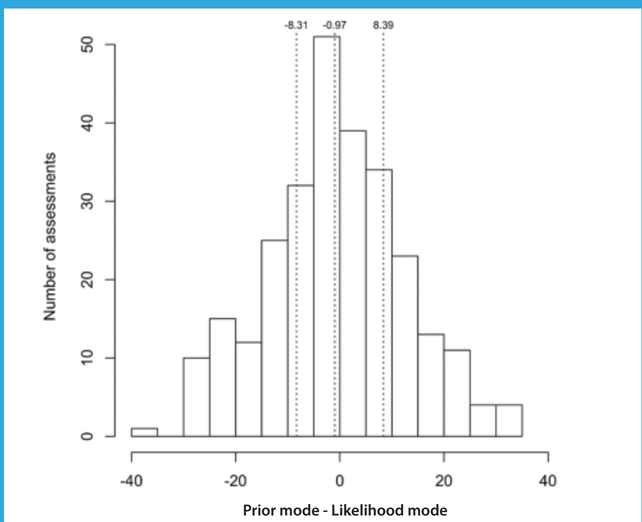
Figure 2 shows the scatterplot of prior modes against likelihood modes. Prior modes and likelihood modes were positively associated with each other. Pearson's correlation coefficient was r = 0.73 (95% CI = 0.67; 0.78). This is very strong evidence ($p < 0.0001$) against the null hypothesis that prior modes and likelihood modes are not associated with each other. The slope of the regression line was $\beta = 0.55$ (95% CI = 0.49; 0.61).

Prior-likelihood conflicts were detected in 20 (7.30%) of the 274 SQUEAC assessments. In 10 SQUEAC assessments with prior-likelihood conflicts the prior mode was below the likelihood mode. In 10 SQUEAC assessments with prior-likelihood conflicts the prior mode was above the likelihood mode (*see Figure 2*). The relative precision of coverage estimates based on likelihood data alone was worse than that which would have been achieved by the EPI-derived gold-standard in seven of the 20 SQUEAC assessments with prior-likelihood conflicts. This means that seven (2.55%) of the 274 SQUEAC assessments were classified as failing due to prior-likelihood conflicts and an inadequate sample size for the likelihood data to estimate coverage with useful precision.

Discussion

The distribution of differences between the prior mode was symmetrical about a central value. This is not consistent with a systematic bias (in either direction) in prior modes. There is a tendency for the prior to overestimate coverage when true coverage is low and to underestimate coverage when true coverage is high. Prior-likelihood conflicts followed this pattern. Prior-likelihood conflicts were detected in 7.30% of the 274 SQUEAC assessments but led to coverage estimates with poor precision in only 2.25% of the 274 SQUEAC assessments. Prior-likelihood conflicts in which the prior mode was below the likelihood mode were equally as common as prior-likelihood conflicts in which the prior mode was above the likelihood mode. These findings indicate that there is no general and systematic failure in SQUEAC. There is an important lesson to be learned from the Isanaka at al. (2018) article. The SQUEAC assessment reported by Isanaka et al. (2018) was not done well. This is admitted in the discussion section of the article. There is no evidence of the use of standard SQUEAC tools and practices such as triangulation by source and method, sampling to redundancy, iteration, the barriers-boosters-questions (BBQ) tool, small studies and surveys, mind-maps, and concept maps. The resolution of conflicting findings by further data collection (iteration) is a key SQUEAC process that was not used. The article states that iteration was not done even when it was indicated. Finding a wide range of candidates for the prior mode, as is reported in the article, should have forced a rethink and further data collection (iteration). The sources for the problematic prior modes were unorthodox. SQUEAC does use caregivers and community members to identify and rank barriers to coverage but these informants are never tasked with responsibility for building the prior. A weak or non-informative prior should always be used with such a wide range of candidates for the prior mode when time and resources for iteration is not available. A key, but understated, finding was that the prior developed by trained

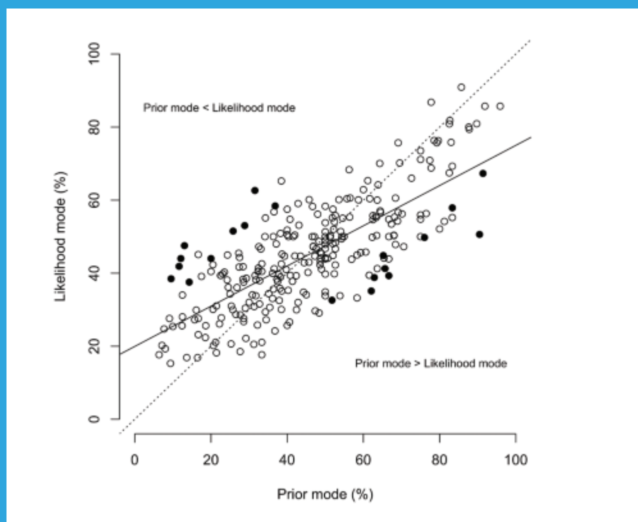
Figure 1 Distribution of the differences between the prior modes and likelihood modes in 274 SQUEAC stage III coverage assessments



Prior and likelihood modes are expressed as percentages.

The dashed vertical lines mark the position of the lower quartile, median, and upper quartile of the differences. Half of all SQUEAC coverage assessments fall between these two lines.

Figure 2 Scatterplot of prior modes against likelihood modes in 274 SQUEAC stage III coverage assessments



The solid line is the ordinary least squares regression line.

The dashed line is the line of equality (i.e. The line on which prior mode = likelihood mode).

Filled circles mark assessments with prior-likelihood conflicts.

staff was unproblematic. The lesson to be learned is that you risk bias when you do SQUEAC with untrained staff, use inappropriate sources, and ignore key SQUEAC processes, methods, and tools.

Conclusion

Prior-likelihood conflicts can and do occur but seldom result in coverage estimates that lack useful precision. They do not lead to a general and systematic overestimation of coverage. The work of Isanaka et al (2018) demonstrates the importance of using trained staff and using SQUEAC processes, methods and tools correctly.

Myatt, M., and Gueverra, E. (2019) Is there a systematic bias in estimates of programme coverage returned by SQUEAC coverage assessments? Field Exchange issue 59, January 2019. www.ennonline.net/fex

References

Isanaka S, Hedt-Gauthier BL, Grais RF, Allen BG. *Estimating program coverage in the treatment of severe acute malnutrition: a comparative analysis of the validity and operational feasibility of two methods*, Population Health Metrics, 2018,16:100,1-9 Full text available from: https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-018-0167-3

Myatt M, Guevarra E, Fieschi L, Norris A, Guerrero S, Schofield L, Jones D, Emru E, Sadler K (2012) *Semi-Quantitative Evaluation of Access and Coverage (SQUEAC) / Simplified Lot Quality Assurance Evaluation of Access and Coverage (SLEAC) Technical Reference,* Food and Nutritional technical Assistance III Project (FANTA-III), FHI 360 / FANTA, Washington, DC, October 2012 www.fantaproject.org/ monitoring-and-evaluation/squeac-sleac

Myatt M (2013) BayesSQUEAC v3.00: *A graphical calculator for Bayesian beta-binomial conjugate analysis of coverage proportions in CTC and CMAM programs using a bootstrap aggregating (bagging) estimator implementing sample size calculation and formal testing for prior-likelihood conflicts*, Brixton Health / Valid International, Oxford, UK, 2013.

Postscript

By Sheila Isanaka, Rebecca F. Grais, and Ben G.S. Allen

We thank Mark Myatt and Ernest Guevarra for adding this important work to the ongoing discussion surrounding appropriate coverage methodologies in the management of acute malnutrition. A 2015 review of coverage methodologies highlighted uncertainty in the use of currently recommended methods for coverage assessment and the need for more peer-reviewed evidence to inform global guidance (Epicentre, 2015). We are delighted that our work (Isanaka et al 2018) may have motivated additional consideration of these important issues and hope that stakeholders and policy makers continue to insist on high-quality, evidence-based experience to inform nutrition programming.

The report by Myatt and Gueverra (2019) shows that coverage estimation using the SQUEAC methodology can yield biased estimations in either direction. For the first time, Myatt and Gueverra quantify the magnitude of this bias, showing only a moderate correlation between the prior and likelihood estimates (Pearson correlation coefficient 95% confidence interval: 0.67 to 0.78). This new evidence is consistent with our findings and supports our conclusion that conflicts between prior and likelihood modes are possible and can lead to biased and imprecise coverage estimates. As discussed by Myatt and Gueverra, the risk of such bias is low when trained staff conduct a survey using appropriate SQUEAC methods. Our work supports this finding, as we similarly show no conflict when an external support team estimated the prior. Our experience, however, goes further than what is possible in the secondary analysis of Myatt and Gueverra to suggest that in resource-limited settings, where sufficient capacity and resources may not be available and the correct methodology may not be faithfully executed, conflict and bias may be more common.

We wholeheartedly agree that coverage assessments should be done using appropriate methods as outlined in the SQUEAC manual (Myatt et al. 2012). Our SQUEAC assessment followed this guidance, including triangulation by source and method, use of the BBQ tool, sampling to redundancy and a small survey. We used booster and barrier weighting by caregivers of severely acutely malnourished (SAM) children as reported in other contexts (Blanárová et al. 2016) as one component in developing the prior mode. The various prior modes in our analyses were combined to simulate different potential scenarios, including situations where external support is not available. This was done for the pedagogic purpose of the study, and as discussed in the paper does not necessarily reflect typical SQUEAC procedures.

We note that the database employed in the analysis of Myatt and Gueverra is likely comprised of surveys conducted by experienced coverage consultants (those provided, for example, by the Coverage Monitoring Network or Valid International) and therefore include prior modes developed by dedicated consultants using gold standard methods that may be less likely to conflict. The analyses further include data from supplementary feeding programmes and does not standardise calculations of coverage estimate according to current guidance to use single coverage (Balegamire S, 2015), analytical choices which may influence the extent to which conflicts and bias were detected. Nonetheless, we welcome the new evidence presented by Myatt and Gueverra (2019) as an important step towards better understanding of the implications of using the SQUEAC methodology for valid coverage estimation.

Overall, we look forward to continued evidence-based and peer-reviewed discussion of appropriate coverage methodologies. Several methodologies are available to monitor programme coverage, and the appropriate study design should be selected in consideration of team capacity, resources and reporting requirements. SQUEAC can be a technically demanding method and requires the appropriate capacity to avoid the potential for bias. As both we and Myatt and Guevarra have shown, conflict and biased coverage estimation are possible and should be considered in selecting the appropriate study design and allocating appropriate resources for assessment.

References

Balegamire S, Siling K, Alvarez Moran JS, Guevarra E, Woodhead S, Norris A, Fieschi L, Paul Binns P, and Myatt M, *A single coverage estimator for use in SQUEAC, SLEAC, and other CMAM coverage assessments*, 2015, Field Exchange 49 Full text available from: https://www.ennonline.net/fex/49/singlecoverage

Blanárová L, Woodhead S, and Myatt M, *Community weighting of barriers and boosters in Democratic Republic of Congo*, 2016, Field Exchange 52 Full text available from: https://www.ennonline.net/fex/52/barriersandboostersdrc

Epicentre. *Open review of coverage methodologies: comments, questions and ways forward*. London: Coverage Monitoring Network; 2015.

Isanaka S, Hedt-Gauthier BL, Grais RF, Allen BG, *Estimating program coverage in the treatment of severe acute malnutrition: a comparative analysis of the validity and operational feasibility of two methods,* Population Health Metrics, 2018,16:100,1-9 Full text available from: https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-018-0167-3

Myatt M, Guevarra E, Fieschi L, Norris A, Guerrero S, Schofield L, Jones D, Emru E, Sadler K (2012) *Semi-Quantitative Evaluation of Access and Coverage (SQUEAC) / Simplified Lot Quality Assurance Evaluation of Access and Coverage (SLEAC) Technical Reference*, Food and Nutritional technical Assistance III Project (FANTA-III), FHI 360 / FANTA, Washington, DC, October 2012 https://www.fantaproject.org/monitoring-and-evaluation/squeac-sleac