Note on the aggregation procedure for SDG indicator 2.5.5

FAO Office of the Chief Statistician, 7 Jan 2019

1. Compilation of SDG indicator 2.5.2

Currently breeds are classified as being "at risk", "not at risk" or with "unknown risk". Unknown includes missing data or data older than 10 years. The frequency of "unknown" is however being included in the denominator of the fraction when computing the indicator.

In general, this practice results in an underestimation of the breeds "at risk", as shown by the following example:

Status	**Frequency**
At risk	n_R
Not at risk	n_{NR}
Unknown	n_U
Total	$n = n_R + n_{NR} + n_U$

The categories of interest are two "At risk" and "Not at risk",

- In the best scenario the breeds with "U" status are all "Not at risk" and therefore:
$$a_R = \frac{n_R}{n}$$
- In the worst scenario breeds with "U" status are all "at risk" and consequently:
$$b_R = \frac{n_R + n_U}{n}$$

In practice, the only valid statement that can be made is that the true percentage of breeds at risk (p_R) is included in the interval:

$$a_R \leq p_R \leq b_R$$

In order to obtain an estimate for p_R, the simplest assumption is that the observed distribution for breeds with known status applies also to the ones with unknown status. Practically, this consists in estimating the fractions of "at risk" only on the subset with known status, i.e.:

$$\hat{p}_R = \frac{n_R}{n_R + n_{NR}}$$

and, as a consequence: $\hat{p}_{NR} = 1 - \hat{p}_R$.

The proportion of breeds of unknown risk status "unknown" status, i.e. $\hat{p}_U = n_U/n$ remains an important indicator that should be computed and published as a quality indicator of the of estimated $\hat{p}_R$ and $\hat{p}_{NR}$; the higher is $\hat{p}_U$ the lower is accuracy of $\hat{p}_R$ and $\hat{p}_{NR}$.

2. Calculation of regional or global aggregates for SDG indicator 2.5.2

According to the established global SDG reporting guidelines, the aggregation of SDG indicators to regional or global levels can only be compiled if a minimum of 50% of countries (in a region for regional reporting or out of all countries worldwide for global reporting) provide up to date information allowing the calculation of the risk status. This will immediately downgrade indicator 2.5.2 from category Tier I to Tier II.

It should be noted that the procedure to be followed to obtain regional or global aggregates for SDG indicator 2.5.2 is to compute a weighted average of the country proportions of the breeds "At risk" and "Not at risk" (which exclude from the denominator the breeds with "unknown" status), with weights given by the number of local breeds within countries.

Formally, if $\hat{p}_{jR}$ is the estimated proportion of "at-risk" local breeds in country *j* then the aggregation formula is:

$$\hat{p}_{+R} = \frac{\sum_{j=1}^{J} \hat{p}_{jR} \times w_j}{\sum_{j=1}^{J} w_j}$$

Where w_j is the weight for country *j* and *J* is the number of reporting countries contributing to the regional aggregate.

It is worth noting that there are two alternative weighting systems:

(i) The weight is the <u>overall</u> number of local breeds in country *j* (including those with unknown risk status), i.e. $w_j = n_j$
The aggregation formula becomes:

$$\hat{p}_{+R} = \frac{\sum_{j=1}^{J} \hat{p}_{jR} \times w_j}{\sum_{j=1}^{J} w_j} = \frac{\sum_{j=1}^{J} \hat{p}_{jR} \times n_j}{\sum_{j=1}^{J} n_j}$$

(ii) The weight is the number of local breed in country *j* with <u>known</u> risk status (excluding those with unknown risk status), i.e. $w_j = n_{jR} + n_{jNR}$.

In this case, the aggregation formula simplifies to:

$$\hat{p}_{+R} = \frac{\sum_{j=1}^{J} \hat{p}_{jR} \times w_j}{\sum_{j=1}^{J} w_j} = \frac{\sum_{j=1}^{J} \hat{p}_{jR} \times (n_{jR} + n_{jNR})}{\sum_{j=1}^{J} (n_{jR} + n_{jNR})} = \frac{\sum_{j=1}^{J} n_{jR}}{\sum_{j=1}^{J} (n_{jR} + n_{jNR})}$$

i.e. it becomes the ratio between the sums of breeds with risk status in the countries and the sum of breeds with known status (excluding those with unknown status)

The aggregation (i) has the advantage of giving more importance to countries with a high number of local breeds. The only drawback is that poor estimates of at-risk status (because of high fraction of breeds has unknown status) will receive a high weight in aggregation if that country has a high number of overall local breeds id compared to other countries. By contrast, although option (ii) is slightly simpler from a computational viewpoint, it will give higher importance to countries having higher number of breeds with known risk status.