Supplementary file 1. Statistical estimation methodology

S1A. Generation of uncertainty intervals on estimated prevalence

The 95% uncertainty or confidence intervals were generated that account for (a) variability in prevalence observed in the data, and (b) modelling error. For each study, we simulated its test-adjusted prevalence, p (in the range 0 to 1), using a beta distribution and adding an error term to the parameters of that beta distribution. Specifically, we assumed that prevalence data of surveys (or other data points) of type $s = 1 \dots S$ were collected (with $S = 1$ for syphilis; and $S \geq 1$ for gonorrhoea) and, for each type s, we assumed that prevalence data from L_s surveys, $p_{ls}, l = 1 \dots L_s$ were available. Let N_{ls} and $n_{ls}, ls = 1 \dots L_s$ denote the sample size and number of tested positive used to estimate p_{ls}. For simplicity, let us assume that the test used to identify infected individuals is perfect. Let $\tilde{P}_{ls}$ be the maximum likelihood estimation of $p_{ls,}\ ls = 1 \dots L_s$. We have $\tilde{P}_{ls} = \frac{n_{ls}}{N_{ls}}$ and $var(\tilde{P}_{ls}) = \frac{\tilde{P}_{ls}(1-\tilde{P}_{ls})}{N_{ls,}}$, and, asymptotically $\tilde{P}_{ls}$ follows a Gaussian distribution. This indicates that, for the purpose of estimating the uncertainty interval, the prevalence p_{ls} can be simulated following a Gaussian distribution of mean $\tilde{P}_{ls}$ and variance $var(\hat{P}_{ls})$. However, if the sample size is not large enough, the asymptotic theory does not hold and simulating p_{ls} using a Gaussian distribution can yield negative prevalence. To overcome this, we opted for a Beta distribution instead. Because we have estimates of the expectation and of the variance of $\tilde{P}_{ls}$, we can estimate the parameters of that beta distribution using the method of moments. The test-adjusted prevalence of each data point p_{ls}, was thus simulated following a beta distribution $P^* \sim \beta(a_{ls}, b_{ls})$ with expectation $E(P^*) = p_{ls}^*$ and variance $var(P^*) = \frac{p_{ls}^*(1-p_{ls}^*)}{N}$, where N is the sample size, the parameter p_{ls}^* is itself simulated using the relation $log\left(\frac{p_{ls}^*}{1-p_{ls}^*}\right) = log\left(\frac{\hat{p}_{ls}}{1-\hat{p}_{ls}}\right) + \varepsilon_{ls}^*$, in which $\hat{p}_{ls}$ is the fitted prevalence (from observed data set) and the error term ε_{ls}^* sampled following a uniform distribution on the set of observed residuals (on the logit scale) with the same (unscaled) weight. Of note, we could have focused on $log(\tilde{P}_{ls})$ and derive its variance using the delta method. However, that technique also relies on the asymptotic theory.

To account for modelling error, we further perturbed the distribution of the simulated prevalence by adding a random noise to the parameters. To avoid further modelling, which may bias our simulations especially if only few data points are available, error terms were randomly selected from the (observed) residuals.

In short, our algorithm for uncertainty analysis for the prevalence can be summarized as follows.

Let:

$$\hat{\varepsilon}_{ls} = log\left(\frac{\tilde{p}_{ls}}{1-\tilde{p}_{ls}}\right) - log\left(\frac{\hat{p}_{ls}}{1-\hat{p}_{ls}}\right), \qquad l_s = 1 \dots L_s, s = 1 \dots S$$

where $\tilde{p}_{ls}$ is the observed prevalence and $\hat{p}_{ls}$ is the estimated prevalence (from the logistic regression for syphilis, or the moving-average estimation for gonorrhoea).

For each data point $l_s = 1 \dots L_s, s = 1 \dots S$:

1- Sample the error term ε^*_{ls} with equal probabilities from the discrete set $\{\hat{\varepsilon}_{ls},\ ls = 1 \dots L_s\}$, .

2- Set $p^*_{ls} = \frac{\hat{p}_{ls}e^{\varepsilon^*_{ls}}}{1-\hat{p}_{ls}+\hat{p}_{ls}e^{\varepsilon^*_{ls}}}$, solution to the equation $log\left(\frac{p^*_{ls}}{1-p^*_{ls}}\right) = log\left(\frac{\hat{p}_{ls}}{1-\hat{p}_{ls}}\right) + \varepsilon^*_{ls}$.

3- Find a_{ls} and b_{ls} such that the mean and variance of the beta distribution with parameters a_{ls} and b_{ls} are p^*_{ls} and $\frac{p^*_{ls}(1-p^*_{ls})}{N_{ls}}$, respectively.

4- Simulate the observed prevalence l_s following the Beta distribution $P^* \sim \beta(a_{ls}, b_{ls})$.

We used the same seed for each bootstrapping iteration, so that prevalence estimates were correlated. Then, the prevalence trend estimation was applied to bootstrapped data, using the same weights as in the initial analysis. This process was repeated 10,000 times. The 2.5% and 95% percentiles of bootstrapped prevalence were used as 95% confidence intervals (CI), and the median bootstrapped prevalence was considered our best estimate.

S1B. Moving average prevalence estimation for gonorrhoea

Time trends in gonorrhoea prevalence were fitted on available prevalence survey data as a moving-average using the formula:

$$\hat{p}_t = \frac{\hat{e}_t}{\hat{E}_t} \qquad Eq1$$

Where:

- $\hat{e}_t = \sum_{s=t_0}^{t_{max}} \sum_{j=1}^{J_S} \tilde{p}_{sj} w_{sj} v_{sj}$, $\hat{E}_t = \sum_{s=t_0}^{t_{max}} \sum_{j=1}^{J_S} w_{sj} v_{sj}$, t_0 is the earliest year with available prevalence, t_{max} is the year with the latest prevalence survey, J_S is the number of surveys conducted during year s, $\tilde{p}_{sj}$ is the (corrected/converted) prevalence, v_{sj} are regional weights

obtained as indicated above; w_{sj} are temporal weights given by: $w_{sj} = \exp\big(|t - s|\log(d)\big)$.

- 'd' is an annual dilution factor that weights the contribution of each data point to the estimation for other years than the study year (earlier and later within the period from first to last data point) downward by a fixed proportion for each additional year away from the estimated year. The dilution factor was set at 20% for both countries.

For countries with qualifying UD or gonorrhoea case report data, these were then added to the estimation, after 'anchoring' them to an average prevalence as initially estimated from prevalence studies (from the country, neighbours and regional estimates). To this end, the case reporting series was converted into a corresponding prevalence trend by multiplying all case reporting rates by a factor such that the Euclidian distance between the resulting series-averaged prevalence and the initially estimated prevalence over the case reporting period was minimal. The resulting case-reporting-trend-based prevalence data series was then added to the dataset of prevalence surveys, and these were used together to estimate a final prevalence time series (without distinction between the two data types), giving all prevalence survey or study data points the weights described above (100% for country, 10% for neighbour and 1% for regional estimate) and each case-reporting-based prevalence proxy a 25% weight.

For the 95% CI estimation, regional estimates [1, 2] were given a notional sample size, inferred from the published 95% confidence bounds (paragraph S1C below). Of note, the 95% CIs thus reflect each study's sample size, whereas the point estimates and the time trend use each study according to its weight (depending on geographical relevance or proximity to the country), but irrespective of sample size. We used the median of bootstrapped prevalence as the final estimate.

S1C. Sample sizes inferred for regional gonorrhoea prevalence estimates

Regional prevalence figures, from the WHO[1] and a meta-analysis of STI prevalences among women attending ANC in sub-Saharan Africa [2], were used to supplement country and neighbouring-country data when estimating the prevalence of gonorrhoea.

Since regional estimates do not have a sample size, an approximate/notional sample size was determined as:

$$\frac{4\times(1.96)^2 p(1-p)}{(U-L)^2}$$

where p is the estimated prevalence, and U and L are the upper and lower bounds of the reported 95% confidence interval. This estimation approach assumes that the infection status of individuals follows a binomial distribution and that the normal approximation to the binomial distribution is sufficiently valid.

Table S1C. Notional sample sizes inferred for regional gonorrhoea prevalence estimates in adult women.

Region & pilot country for which this regional estimate was used	***Year***	***Regional prevalence estimate (95% CI)***	***Inferred notional sample size***
Southern Africa – used for Zimbabwe	1996	3.7% (2.7-4.6%) [2]	1,517
Southern Africa – used for Zimbabwe	2005	3.8% (2.28-6.04%) [2]	192
Southern Africa – used for Zimbabwe	2012	4.9% (1.8-7.9%) [1]	394
Eastern Mediterranean – used for Morocco	2012	0.62% (0.41-0.83%) [1]	5,444

References for Supplementary file 1:

1. Newman L, Rowley J, VanderHoorn S, et al. Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012. *PLoS One*. 2015;**10**:e0143304.
2. Chico RM, Mayaud P, Ariti C, et al. Prevalence of malaria and sexually transmitted and reproductive tract infections in pregnancy in sub-Saharan Africa: a systematic review. *JAMA*. 2012;**307**:2079-86.