Variance estimation for effective coverage measures: A simulation study

Sara M Sauer[1], Thomas Pullum[2], Wenjuan Wang[2,3]*, Lindsay Mallick[4], Hannah H Leslie[5]

[1] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

[2] The Demographic and Health Surveys (DHS) Program, Division of International Health and Development, ICF, Rockville, Maryland, USA

[3] Division of AIDS, Behavioral, and Population Sciences; Center for Scientific Review, National Institutes of Health; Bethesda, Maryland, USA

[4] The DHS Program, Avenir Health; Glastonbury, Connecticut, USA

[5] Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

* This work was prepared while Wenjuan Wang was employed at ICF. The opinions expressed in this article are the author's own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Correspondence to:

Hannah H. Leslie, PhD, MPH
90 Smith Street, suite 326
Boston, Massachusetts 02120
USA
hleslie@hsph.harvard.edu

Background Effective coverage research is increasing rapidly in global health and development, as researchers use a range of measures and combine data sources to adjust coverage for the quality of services received. However, most estimates of effective coverage that combine data sources are reported only as point estimates, which may be due to the challenge of calculating the variance for a composite measure. In this paper, we evaluate three methods to quantify the uncertainty in the estimation of effective coverage.

Methods We conducted a simulation study to evaluate the performance of the exact, delta, and parametric bootstrap methods for constructing confidence intervals around point estimates that are calculated from combined data on coverage and quality. We assessed performance by computing the number of nominally 95% confidence intervals that contain the truth for a range of coverage and quality values and data source sample sizes. To illustrate these approaches, we applied the delta and exact methods to estimates of adjusted coverage of antenatal care (ANC) in Senegal. We used household survey data for coverage and health facility assessments for readiness to provide services.

Results With small sample sizes, when the true effective coverage value was close to the boundaries 0 or 1, the exact and parametric bootstrap methods resulted in substantial over or undercoverage and, for the exact method, a high proportion of invalid confidence intervals, while the delta method yielded modest overcoverage. The proportion of confidence intervals containing the truth in all three methods approached the intended 95% with larger sample sizes and as the true effective coverage value moved away from the 0 or 1 boundary. Confidence intervals for adjusted ANC in Senegal were largely overlapping across the delta and exact methods, although at the sub-national level, the exact method produced invalid confidence intervals for estimates near 0 or 1. We provide the code to implement these methods.

Conclusions The uncertainty around an effective coverage estimate can be characterized; this should become standard practice if effective coverage estimates are to become part of national and global health monitoring. The delta method approach outperformed the other methods in this study; we recommend its use for appropriate inference from effective coverage estimates that combine data sources, particularly when either sample size is small. When used for estimates created from facility type or regional strata, these methods require assumptions of independence that must be considered in each example.

Progress towards global development goals, including Sustainable Development Goal (SDG) 3: good health and well-being, is dependent on the operation of complex systems in both high and low-resource settings [1]. Achieving the ambitious health targets in Goal 3 demands that health systems, in order to successfully manage population health, must prevent disease, treat acute and chronic illness, and respond to emergencies and disasters [2].

While global monitoring early in the 2000s focused on health system coverage and the proportion of those in need who receive a service, there is increasing recognition that monitoring in the SDG era must evaluate the effectiveness of service in order to capture the true value of health systems [3,4]. This is reflected in the growing research on effective coverage of health interventions [5]. Effective coverage can be defined as the fraction of potential health gain successfully delivered [3], or more operationally, as the "people who need health services obtain[ing] them in a timely manner and at a level of quality necessary to obtain the desired effect and potential health gains" [6].

Effective coverage research is increasing rapidly in global health, as researchers use a range of measures to adjust coverage for the quality and effects of care received [5]. The calculation of measures of effective coverage frequently requires combining data from different sources, such as administrative records that provide service coverage and clinic data for the content or quality of services [7]. A recent review and framework for effective coverage recommended reserving the term "effective coverage" for outcome-adjusted coverage and specifying input-adjusted or process-adjusted coverage based on the type of measure used for service delivery [5]. Here we use "effective coverage" generally for any adjusted coverage estimate, given its use in existing literature, and refer to specific estimates as suggested in this framework.

As an example, multiple studies have attempted to quantify quality-adjusted coverage of antenatal care (ANC). All pregnant women are encouraged to attend formal health care during pregnancy for assessment, preventive measures, and counseling. The focused ANC model recommended by the World Health Organization (WHO) in 2002 included four visits for basic care, with the first visit taking place ideally within the first 12 weeks of gestation [8]. The Millennium Development Goal to improve maternal health included attendance at four ANC visits as a key target for service coverage [9]. In 2016, the WHO released guidelines that increased the recommended contacts to eight visits [10], although this has not yet been widely adopted as a measure of coverage. Studies on quality-adjusted coverage have used both one visit and four visits to quantify ANC coverage, which is based typically on maternal self-report in household surveys [11-19]. Facility measures have similarly followed WHO guidance on the expectations for ANC and have generally focused on inputs to care, with studies that define structural quality measures for health facilities such as functional diagnostics and medications [11,15,16]. Across all studies, despite the variation in measurement, crude coverage of ANC substantially exceeded the adjusted coverage. Failing to consider either facility readiness or quality of care can dramatically overstate health system performance.

Multinational and academic groups have called for a transition from coverage measures to effective coverage measures in order to appropriately benchmark and monitor national progress towards the SDGs [2,4,20]. Effective coverage metrics require accurate estimation of uncertainty around each estimate. While it is routine to report variance for effective coverage estimates that use one data source [19,21,22], most research that integrates multiple data sources to estimate effective coverage provides only point estimates. A 2018 literature review [5] found that of six papers that combined population and health facility data sources, four did not report an estimate of the variance of the point estimate [11,12,15,16]; one used the exact method for the product of two variables [17], and the sixth used a Taylor series expansion [19]. Recently published work on the effective coverage of childbirth defined and employed the delta method for variance calculation [23,24].

This paper aims to provide a formal definition of effective coverage; describe the exact, delta, and bootstrap methods of constructing confidence intervals around effective coverage estimates; test the methods' performance with a simulation study; demonstrate their application to the estimation of effective coverage of ANC in Senegal; and provide guidance for applied research.

METHODS

Effective coverage can be defined at three levels: *first*, for a specific combination of facility type f and region r; *second*, for region r, combining all facility types; *third*, at the national level, combining all regions. When combining two data sources without an exact link between individuals and health facilities, incorporating health facility type improves the accuracy of the overall estimates [25]. Here we present the

analytical approaches for all three levels. We conducted the simulations at a single level to demonstrate method performance in the most simple setting.

Calculating effective coverage

Estimates of effective coverage are specific to a particular type of intervention. For example, an intervention could be ANC for recently pregnant women age 15-49, and the relevant readiness measure could be a binary or continuous measure of facility preparedness to offer high-quality ANC. For illustration, we define a binary random variable X_{rf} that represents crude coverage and a binary random variable Y_{rf} that represents facility readiness, both of which are specific to sub-national region r and facility type f. (The full derivation of effective coverage is shown in Appendix S1 of the **Online Supplementary Document**). The distributions of these random variables are indexed by a mean parameter $P(X_{rf}=1)=P_{xrf}$ and $P(Y_{rf}=1)=P_{yrf}$, respectively. Since P_{xrf} and P_{yrf} are usually unknown in practice, these probabilities are estimated with sample means. Let p_{xrf} be the estimated measure of coverage, such as the sample proportion of women who seek ANC in facilities of type f in region r. Let p_{yrf} be the estimated measure of readiness, such as the sample proportion of facilities of type f in region r that satisfy minimal criteria for ANC. The estimated effective coverage p_{rf} for this combination of r and f is a product:

$$p_{rf}=p_{xrf}\times p_{yrf} \tag{1}$$

At the regional level, the effective coverage p_r is obtained by adding the contributions from all facilities:

$$p_r=\Sigma_f\times p_{rf} \tag{2}$$

This can also be viewed as a weighted sum of readiness p_{yrf} where the weights are the proportion of women attending that facility type (p_{xrf}, $\Sigma_f\times p_{rf}=1$). The national effective coverage p is a weighted average of the regional values. The weights w_r ($\Sigma_r\times w_r=1$) are proportional to the relevant denominator for each region, such as the number of women or the number of pregnancies:

$$P=\Sigma_r\times w_r p_r \tag{3}$$

Calculating a confidence interval for effective coverage

We describe three methods for constructing a 95% confidence interval around an effective coverage estimate: two analytical approaches, the *exact* and the *delta* methods, and a computer-based approach, the *bootstrap* method, which involves repeated sampling with parameters estimated from the observed data. The derivation for confidence interval construction using the *delta* method was presented previously in Wang et al. [23,26].

Method 1: The exact method

The exact method of calculating the sampling variance of effective coverage estimates is based on a formula for the exact variance of the product of two independent random variables derived by Goodman in 1960 [27]. By definition, the variance of the product of two random variables Z and M is $Var(ZM)=E[(ZM)^2]-(E[ZM])^2$. If the two variables are independent, then $E[Z^2M^2]=E[Z^2]E[M^2]$ and $E[ZM]=E[Z]E[M]$, from which it follows that $Var(ZM)=E[Z^2]E[M^2]-(E[Z]E[M])^2$. Since $E[Z^2]=Var(Z)+E[Z]^2$, and $E[M^2]=Var(M)+E[M]^2$, $Var(ZM)$ can be re-expressed as

$$Var(ZM)=(Var(Z)+E[Z]^2)\times(Var(M)+E[M]^2)-(E[Z]E[M])^2 \tag{4}$$

This formula can be used to calculate the variance of $p_{rf}=p_{xrf}\times p_{yrf}$ denoted by $se^2(p_{rf})$:

$$se^2(p_{rf})=[Var(p_{xrf})+E[p_{xrf}]^2][Var(p_{yrf})+E[p_{yrf}]^2]-E[p_{rf}]^2 \tag{5}$$

Assuming a normal approximation to the distributions of p_{xrf} and p_{xrf} yields $Var(p_{xrf})=(P_{xrf}(1-P_{xrf}))/n_{xrf}$ and $Var(p_{yrf})=(P_{yrf}(1-P_{yrf}))/n_{yrf}$, where n_{xrf} and n_{yrf} are the sample sizes if the samples were collected via simple random sampling, or the effective sample sizes, if the samples were collected through a complex survey design. These quantities can be estimated using p_{xrf} and p_{yrf} in place of P_{xrf} and P_{yrf}, respectively. The *estimated* variance of $p_{rf}=p_{xrf}p_{yrf}$, denoted by $s^2(p_{rf})$, is then given by:

$$s^2(p_{rf})=[(p_{xrf}(1-p_{xrf}))/n_{xrf}+p^2_{xrf}][(p_{yrf}(1-p_{yrf}))/n_{yrf}+p^2_{yrf}]-p^2_{rf} \tag{6}$$

The approach under the exact method is to calculate the estimated standard error $s_{rf}=s(p_{rf})$ using (6), and then use a normal approximation to the distribution of p_{rf} to calculate

$$L_{prf}=p_{rf}-1.96s_{rf} \text{ and } U_{prf}=p_{rf}+1.96s_{rf} \tag{7}$$

as the lower and upper ends, respectively, of the 95% confidence interval for effective coverage.

To summarize, this method is exact in equations (4) and (5); equation (6) provides an estimate of the variance based on observed sample proportions; and (7) gives a symmetric Wald-type 95% confidence interval. When p_{rf} is close to either 0 or 1, it is possible for this confidence interval to cross the 0,1 boundary and yield an invalid confidence interval.

Expression (6) can be modified to take into account the facility and regional levels as follows:

$$s^2{}_{rf} = [(p_{xrf}(1-p_{xrf}))/n_{xrf} + p^2{}_{xrf}] \, [(p_{yrf}(1-p_{yrf}))/n_{yrf} + p^2{}_{yrf}] - p^2{}_{rf}$$

$$s^2{}_{r} = \Sigma_f[(p_{xrf}(1-p_{xrf}))/n_{xrf} + p^2{}_{xrf}] \, [(p_{yrf}(1-p_{yrf}))/n_{yrf} + p^2{}_{yrf}] - p^2{}_{rf}$$

$$s^2 = \Sigma_f w^2{}_r \Sigma_f \, [(p_{xrf}(1-p_{xrf}))/n_{xrf} + p^2{}_{xrf}] \, [(p_{yrf} \,(1-p_{yrf}))/n_{yrf} + p^2{}_{yrf}] - p^2{}_{rf}$$

Note that we assume independence (covariance = 0) between the effective coverage estimates for different facility/regional levels in the expressions above (as well as in the following methods). This is an assumption that should be evaluated in each application of these methods.

Method 2: The delta method

An alternative approach to constructing a 95% confidence interval for a proportion is to transform the proportion to a logit scale, calculate the confidence interval on that scale, and then take antilogits of the confidence interval endpoints to return to the original scale. Specifically, consider the proportion P, which is estimated by p. Define

$$F = logit(p) = b \tag{8}$$

In this representation, b is the intercept or constant term from a logit regression with no covariates. Let s_b be the standard error of b. Under the assumption that b, rather than p, has a normal sampling distribution, a 95% confidence interval for the population value of b with lower and upper limits L_F and U_F, respectively, is calculated as

$$L_F = F - 1.96s_b \text{ and } U_F = F + 1.96s_b \tag{9}$$

We calculate the inverse transformation of the logit, or the antilogit, to obtain the lower and upper limits of the 95% confidence interval for P:

$$L = \exp(L_F)/[1 + \exp(L_F)] \text{ and } U = \exp(U_F)/[1 + \exp(U_F)] \tag{10}$$

If n_0 and n_1 are the number of cases with $X = 0$ and $X = 1$, respectively, then $b = \log(n_1/n_0)$ and $s_b = \text{sqrt}[(1/n_0) + (1/n_1)]$. Standard statistical software will produce b, s_b, L_F, and U_F, and can adjust for sample weights, clustering, and stratification, if applicable.

When p represents the effective coverage estimate, we cannot use an exact formula to calculate s_b, the standard error of $logit(p)$, because $logit(p)$ cannot be factored into a product of two variables. We can, however, calculate the b's and s's from the readiness and coverage data separately, and use these to approximate s_b. Suppose that b_x and b_y are the intercepts from logit regressions of X and Y separately, with sampling $s^2{}_{bx}$ and $s^2{}_{by}$, respectively, all of which are produced with standard statistical packages. If $F = F(b_x, b_y)$ is a joint function of b_x and b_y, then an approximation to the sampling variance of F is given by

$$s^2 = \left(\frac{\partial F}{\partial b_x}\right)^2 s^2_{bx} + \left(\frac{\partial F}{\partial b_y}\right)^2 s^2_{by} = \sum_{k=x}^{y}\left(\frac{\partial F}{\partial b_k}\right)^2 s^2_{bk} \tag{11}$$

This well-known approximation is sometimes described as the delta method [28].

Now we apply the delta method to the logit of effective coverage. The logit transformation is applied to three levels of effective coverage: *first*, for facility type f in region r; *second*, for region r; and *third*, at the national level, for which no subscripts are needed. The transformations are

$$F_{rf} = logit(p_{rf}) \tag{12}$$

$$F_r = logit(p_r) \tag{13}$$

$$F = logit(p) \tag{14}$$

At these three levels, (11) takes the following forms:

$$s^2_{brf} = \sum_{k=y}^{x}\left(\frac{\partial F_{rf}}{\partial b_{krf}}\right)^2 s^2_{bkrf} \tag{15}$$

$$s_{br}^2 = \sum_f \sum_{k=y}^x \left(\frac{\partial F_r}{\partial b_{bkrf}} \right)^2 s_{bkrf}^2 \tag{16}$$

$$s_b^2 = \sum_r \sum_f \sum_{k=y}^x \left(\frac{\partial F}{\partial b_{krf}} \right)^2 s_{bkrf}^2 \tag{17}.$$

The components for the chained partial derivatives are provided in Appendix S2 of the **Online Supplementary Document**. When the components are multiplied together as described, the partial derivatives required for equations (15), (16), and (17) are given by (18), (19), and (20), respectively:

$$\frac{\partial F_{rf}}{\partial b_{krf}} = \left(1 - p_{krf}\right) / \left(1 - p_{rf}\right) \tag{18}$$

$$\frac{\partial F_r}{\partial b_{krf}} = p_{rf}\left(1 - p_{krf}\right) / \left[p_r \left(1 - p_r\right) \right] \tag{19}$$

$$\frac{\partial F}{\partial b_{krf}} = w_r p_{rf}\left(1 - p_{krf}\right) / \left[p \left(1 - p\right) \right] \tag{20}$$

	Delta method: $antilogit(F \pm 1.96 \times s_{b_{rf}})$, where $s_{b_{rf}}^2 =$	Exact method: $p \pm 1.96 s_{rf}$, where $s_{rf}^2 =$
Binary coverage indicator / Binary readiness indicator	$\sum_{k=y}^x \left(\frac{\partial F_{rf}}{\partial b_{krf}} \right)^2 s_{b_{krf}}^2$, with $\frac{\partial F_{rf}}{\partial b_{krf}} = (1 - p_{krf})/(1 - p_{rf})$	$[(p_{xrf}(1 - p_{xrf}))/n_{xrf}$ $+\ p_{xrf}^2\ ][(p_{yrf}(1 - p_{yrf}))/n_{yrf}$ $+\ p_{yrf}^2] - p_{rf}^2$
Binary coverage indicator / Continuous readiness score	$\sum_{k=y}^x \left(\frac{\partial F_{rf}}{\partial b_{krf}} \right)^2 s_{b_{krf}}^2$, with $\frac{\partial F_{rf}}{\partial b_{yrf}} = 1/[p_{yrf}(1 - p_{rf})]$ and $\frac{\partial F_{rf}}{\partial b_{xrf}} = (1 - p_{xrf})/(1 - p_{rf})$	$[(p_{xrf}(1 - p_{xrf}))/n_{xrf}$ $+\ p_{xrf}^2\ ][[\frac{1}{n_{yrf}}(\frac{\sum_i (Y_{irf} - \overline{Y}_{rf})^2}{(n_{yrf} - 1)})]$ $+\ p_{yrf}^2\ ] - p_{rf}^2$

Figure 1. Delta and exact methods for construction of a 95% confidence interval around the effective coverage estimate for a specific facility type f in a specific region r, p_{rf}.

To summarize this approach, partial derivatives are calculated from (18) or (19) or (20). The s's are produced by logit regression software. At each level, a confidence interval is constructed for F, and the endpoints of that interval are converted to a confidence interval (L, U) for P by taking antilogits, as in (10).

In some cases, facility readiness may be a continuous score rather than a binary measure. **Figure 1** summarizes the variance calculations for the exact and delta methods in the case of a binary coverage indicator and a binary or a continuous readiness measure. Details of the modifications to the formulas in the case of a continuous readiness measure are shown in Appendix S3 of the **Online Supplementary Document**.

Method 3: The bootstrap method

The final approach we consider is a bootstrap method, a computer-based approach that involves repeated sampling from a distribution or the observed data [29]. Bootstrapping methods are used increasingly to generate variance estimates for complex parameters. We first compared a non-parametric bootstrap in which observations are re-sampled from observed data (with replacement) to a parametric bootstrap in which observations are generated with parameters estimated from the original sample. Initial comparisons revealed consistent undercoverage when using a nonparametric bootstrap. We therefore focus on the parametric bootstrap for the full analysis. When coverage and readiness are proportions as presented above, the steps for constructing a 95% confidence interval for the effective coverage estimate for a particular region r and facility type f based on a simple random sample for each data source are as follows:

- Obtain p_{xrf} and p_{yrf} from the observed data,
- Use these estimates to generate B samples from Binomial(n_{xrf}, p_{xrf}), and B samples from Binomial(n_{yrf}, p_{yrf}),
- Divide the sampled values by n_{xrf}, and n_{yrf} respectively to get the B-vectors p_{xrf_boot} and p_{yrf_boot},
- Multiply the corresponding elements of p_{xrf_boot} and p_{yrf_boot} to get the B-vector p_{boot},
- Obtain the upper and lower limits of the 95% confidence interval using the 2.5th and 97.5th percentiles of the distribution of p_{boot},

If readiness is a continuous score, this procedure can be modified to generate data using an appropriate distribution.

Simulation study

We designed a simulation study to test the performance of the confidence interval construction methods in settings with small and modest sample sizes. In many cases in applied research, the sample size for health facilities will be much smaller than the individual sample. We tested settings of both equal and unequal sample sizes.

We assume that the crude coverage of a service in the population is a binary measure with probability P_{xrf} and health system quality is a binary measure with probability P_{yrf}. We test combinations of P_{xrf} and P_{yrf} over the range (0.02, 0.98) and with sample sizes for n_{xrf}, the number of individuals sampled, and n_{yrf}, the number of facilities sampled, that varied between 50 and 500 (all settings considered are presented in Table 1). In each case, the sample of individuals and the sample of facilities are simple random samples. As a further analysis, we consider the case in which the service readiness measure is a continuous score on 0,1.

Table 1. Simulation settings

Term	Definition	Simulation settings
P_{xrf}	Proportion of individuals in need accessing service (crude coverage)	(0.02, 0.04, 0.06, 0.08, 0.1, 0.15, 0.2, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.6, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.92, 0.94, 0.96, 0.98)
P_{yrf}	Proportion of health facilities rated as 'high-quality'	(0.02, 0.04, 0.06, 0.08, 0.1, 0.15, 0.2, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.6, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.92, 0.94, 0.96, 0.98)
n_{xrf}	Number of individuals sampled	$(n_{xrf}, n_{yrf}) = (50, 50), (100, 100), (200, 200), (300, 300), (400, 400),$
n_{yrf}	Number of facilities sampled	$(500, 500), (100, 50), (200, 50), (300, 50), (400, 50), (500, 50)$

For each simulation setting, we generated 10 000 data sets. For each simulated data set, we computed the effective coverage estimate $p_{rf} = p_{xrf}p_{yrf}$ and constructed 95% confidence intervals around p_{rf} using the exact, delta, and bootstrap methods. For the bootstrap method, 10 000 bootstrap samples were taken to construct the confidence interval at each iteration. Across the 10 000 simulated iterations, we calculated the estimated coverage probabilities of the three methods for constructing a 95% confidence interval by computing the proportion of iterations in which the true effective coverage value was contained in the 95% confidence interval. When computing the estimated coverage probabilities for the exact and bootstrap methods, we removed the iterations that resulted in a degenerate confidence interval, which occurs when $p_{rf} = 0$ or 1. When estimating the coverage probabilities for the delta method, we removed the iterations that resulted in an undefined confidence interval, which occurs when when $p_{rf} = 0$ or when either p_{xrf} or $p_{yrf} = 1$. We recorded the number of undefined confidence intervals for the delta method, which is the maximum number of iterations removed across the three methods, and the number of invalid estimates for the exact method (the instances in which the confidence interval bounds fall below 0 or above 1).

Applied example

We demonstrate the application of these methods using real data on ANC coverage and facility readiness in the Republic of Senegal. The West African nation of Senegal is home to over 13 million individuals who live in 14 administrative regions [30]. Senegal is classified as a low-income country by the World Bank. The public health sector is organized into hospitals at the top, followed by intermediate health centers and peripheral health posts. Health extension huts (*cases de santé*) offer limited services, which are largely for reproductive, maternal, and child health, while the private sector also provides health services. From 2012 to 2017, the National Agency for Statistics and Demography partnered with ICF International to implement a simultaneous and continuous Demographic and Health Survey (DHS) and Service Provision Assessment (SPA). Each assessment was administered annually over the 5-year period. We relied on the 2017 surveys, using the DHS to define coverage and the SPA to define readiness.

The DHS is a population-based household sample, in which women age 15 to 49 answer questions about their reproductive history. We defined ANC coverage based on responses from women with a live birth in the past 2 years, and calculated three measures based on global and national standards of care:

- Coverage 1: self-report of any ANC with a formal health care provider
- Coverage 4: self-report of 4 or more ANC visits
- Coverage 8: self-report of 8 or more ANC visits

The SPA is a health facility assessment that includes an audit of service availability and readiness, as well as provider interviews and observations of care. We use the facility audit to define readiness to provide high-quality ANC, which was limited to health facilities that provide ANC. To best illustrate the variance estimation methods, we defined two binary indicators (one prevalent, one rare) and a bounded score as quality measures:

- Readiness 1: Availability of manual or digital blood pressure apparatus (binary)

- Readiness 2: Availability of hemoglobin and urine protein diagnostic capacity (binary)

- Readiness 3: ANC service readiness, the proportion of essential inputs in place across four domains as defined by the WHO: basic amenities (guidelines, visual aids, and provider training); equipment (manual or digital blood pressure apparatus); diagnostics (urine and anemia testing capacity); and medication (iron and folic acid, tetanus toxoid, and IPTp and ITNs for malaria) [31]. We weighted items evenly within domain and domains equally within the score.

To define adjusted coverage, we calculated service readiness by strata defined by the fourteen regions and five health facility types (government hospital, government health center, government health clinic/post, private hospital/health center/clinic, health hut). Coverage was estimated for the same strata. In this example, we assigned women who identified multiple sources of care the source with the highest readiness and assumed that all visits took place at that source or a similar source. This will overestimate adjusted coverage for women using multiple sources of care of varying readiness. We calculated adjusted coverage as the product of coverage and readiness within strata and estimated national adjusted coverage as the average of stratum-specific estimates weighted by population size (women reporting live births in the previous 2 years).

We repeated this procedure for the nine combinations of the coverage and quality measures and used the exact and delta methods to quantify variance in each case. The complex sample designs used for both DHS and SPA are accounted for by using the survey setting. For the DHS, the strata are based on sub-national region and urban/rural residence; the cluster is the primary sampling unit, and individual weights are the women's sampling weight that account for non-response and sampling probability. For the SPA, the strata are based on sub-national region, health facility type, and managing authority; the primary sampling unit is the facility, and the individual weights are the facility sampling weight. Although the bootstrap method, particularly the non-parametric bootstrap, can be modified for complex survey data [32], expanding the parametric bootstrap method for a parameter derived from two complex surveys is beyond the scope of this paper and is an area for future research.

Simulations were conducted in R (The R Foundation for Statistical Computing, Vienna, Austria). Applied analyses were conducted with Stata version 15 (StataCorp, College Station, Texas, USA). A program for calculating point estimates and confidence intervals in Stata for national and sub-national analyses like the applied example as well as sub-sample analyses to assess equity is included in the **Online Supplementary Document** and at this link (https://osf.io/9nsaf/?view_only=681d595548014a17a4a66669 0e708336), as is the simulation code in R.

Ethical approval

The original survey implementers obtained ethical approvals for data collection; the Harvard University Research Protection Program deemed this analysis exempt from human subjects review.

RESULTS

Simulation study

The simulation results for the estimated coverage probabilities are shown in **Figure 2**, **Figure 3**, and **Figure 4** for the exact, delta, and bootstrap methods, respectively. The maximum proportion of simulations with undefined estimates is shown in Appendix S4 in the **Online Supplementary Document**.

When the sample size is small, and when the true proportions P_{xrf} and/or P_{yrf} are close to 0 or 1, the confidence intervals constructed using the exact method result in severe overcoverage for the most extreme values of P_{xrf} and P_{yrf}, which then becomes undercoverage as P_{xrf} and P_{yrf} increase. Of the iterations that do not result in degenerate confidence intervals, the nominally 95% confidence interval captures the truth 100% of the time in the extreme case of $P_{xrf} = 0.02$ and $P_{yrf} = 0.02$ with $(n_{xrf}, n_{yrf}) = (50, 50)$, as shown in
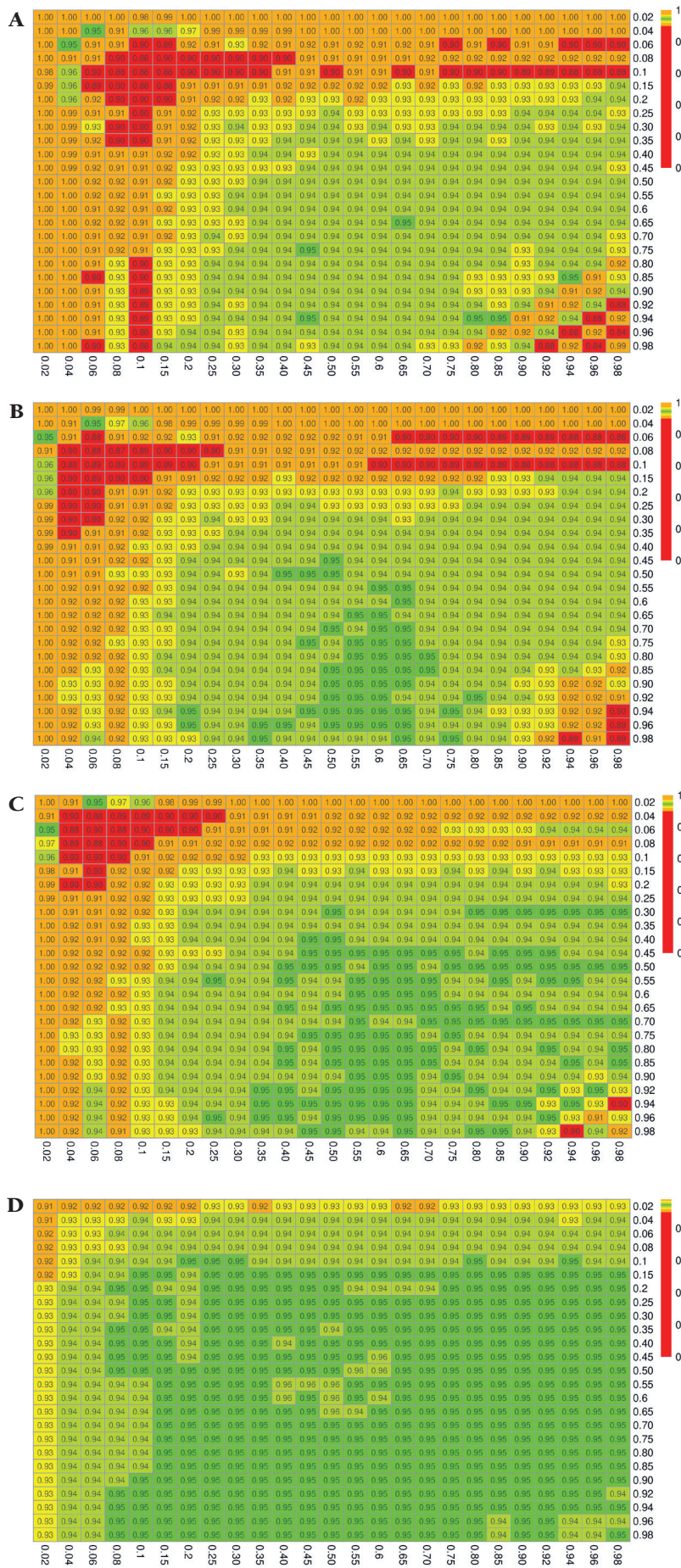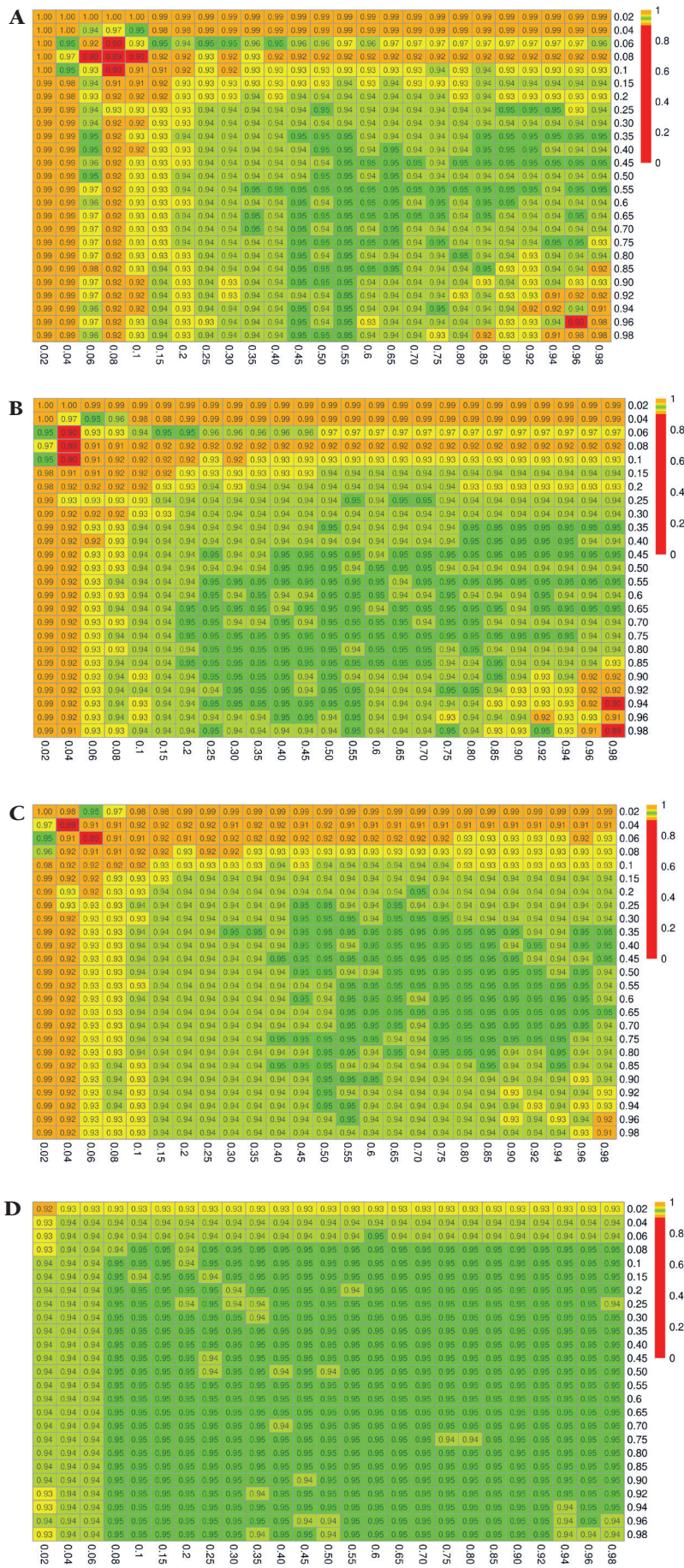
Figure 2. Estimated 95% coverage probabilities for the confidence intervals constructed using the exact method. P_{xrf} and P_{yrf} vary over the range (0.02, 0.98). **Panel A.** $(n_{xrf}, n_{yrf}) = (50, 50)$. **Panel B.** $(n_{xrf}, n_{yrf}) = (100, 50)$. **Panel C.** $(n_{xrf}, n_{yrf}) = (100, 100)$. **Panel D.** $(n_{xrf}, n_{yrf}) = (500, 500)$.

Figure 3. Estimated 95% coverage probabilities for the confidence intervals constructed using the delta method. P_{xrf} and P_{yrf} vary over the range (0.02, 0.98). **Panel A.** $(n_{xrf}, n_{yrf}) = (50, 50)$. **Panel B.** $(n_{xrf}, n_{yrf}) = (100, 50)$. **Panel C.** $(n_{xrf}, n_{yrf}) = (100, 100)$. **Panel D.** $(n_{xrf}, n_{yrf}) = (500, 500)$.

Figure 4. Estimated 95% coverage probabilities for the confidence intervals constructed using the parametric bootstrap method. P_{xrf} and P_{yrf} vary over the range (0.02, 0.98). **Panel A.** $(n_{xrf}, n_{yrf}) = (50, 50)$. **Panel B.** $(n_{xrf}, n_{yrf}) = (100, 50)$. **Panel C.** $(n_{xrf}, n_{yrf}) = (100, 100)$. **Panel D.** $(n_{xrf}, n_{yrf}) = (500, 500)$.

Figure 2, Panel A. Moreover, as shown in Appendix S4 of the **Online Supplementary Document**, all of the non-degenerate confidence intervals constructed in this setting are invalid, with the lower bound falling below 0. For small sample sizes and values of P_{xrf} and P_{yrf} between 0.06 and 0.25, the exact method yielded confidence intervals that contained the truth closer to 90 than 95% of the time. This undercoverage may be due to the use of a normal approximation, which may not be valid with small sample sizes. A similar pattern persists at sample sizes of 100 (**Figure 2**, Panels B and C). When both n_{xrf} and n_{yrf} are 500, the coverage is generally at or close to the nominal level (0.95), except for very extreme values of P_{xrf} or P_{yrf} (**Figure 2**, Panel D).

The confidence intervals constructed with the delta method yield estimated coverage probabilities that are generally close to the nominal level of 95%, although for small sample sizes and extreme values of P_{xrf} or P_{yrf}, this method results in overcoverage as high as 98% (**Figure 3**, Panel A). Coverage is uniformly close to the nominal 95% level once sample sizes reach 400 to 500.

The results for the bootstrap method mirror those described for the exact method, with slightly better performance compared to the exact method. In particular, we see overcoverage (99-100%) for small sample sizes and very extreme values of P_{xrf} or P_{yrf}, which then becomes undercoverage for values of P_{xrf} and P_{yrf} up to approximately 0.25 (**Figure 4**, Panel A). With samples of at least 500, coverage is no lower than 92% (**Figure 4**, Panel D).

Appendix S4 in the **Online Supplementary Document** includes full results with the number of invalid confidence interval bounds for the exact method, which occurred in as high as 100% of the iterations yielding non-degenerate confidence intervals for effective coverage estimates near the boundary, when $n_{xrf} = n_{yrf} = 50$, and occurred as much as 26% of the time when $n_{xrf} = n_{yrf} = 500$.

When using a continuous score as a readiness estimate instead of a binary indicator, the relative performance of the confidence interval construction methods was comparable, with the delta method demonstrating 95% coverage plus or minus only a few percent in nearly all settings (Appendix S5 in the **Online Supplementary Document**, **Figure 3**). The results for the exact and bootstrap method, were similar to the results in the setting with a binary readiness estimate, except that the over- and undercoverage arises in settings with small sample sizes and extreme values of P_{xrf} (binary coverage) only. The exact method and to a lesser extent the bootstrap method resulted in coverage below the nominal 95% level in small samples when P_{xrf} true coverage was less than 10% at any level of readiness or when true coverage was greater than 0.9 and readiness exceeded 0.5 (Appendix S5 in the **Online Supplementary Document**, **Figure 2**, Panel A, **Figure 4**, Panel A). Coverage probabilities were all close to 95% once both sample sizes reached 500.
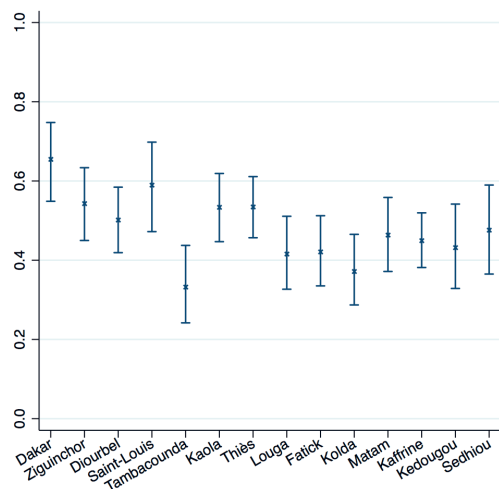
Applied example

Figure 5 shows the adjusted coverage estimates and 95% confidence intervals constructed with the exact and delta methods for the nine combinations of coverage and quality measures, as described in the
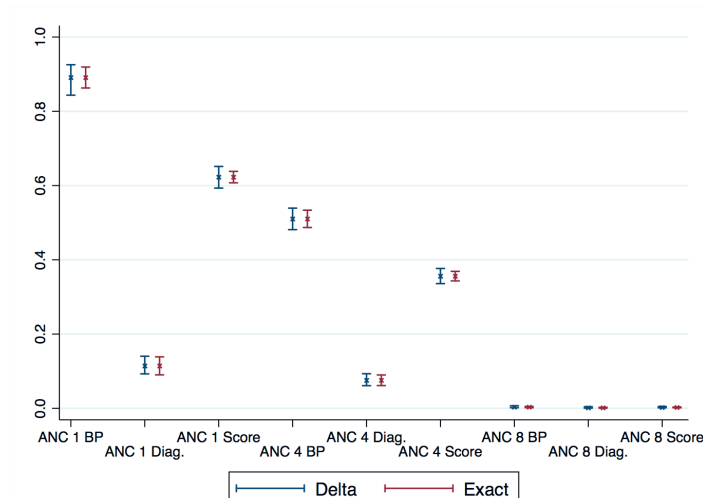
Figure 5. Adjusted coverage estimates of antenatal care in Senegal with 95% confidence intervals using the delta and exact methods. Coverage is defined as 1, 4, or 8 antenatal care visits (ANC 1, ANC 4, ANC 8 respectively). Readiness is defined based on functional blood pressure apparatus (BP), functional diagnostics (Diag.), or summary score (Score).

Methods section. Adjusted coverage of ANC in Senegal differed as expected, based on the definitions of coverage and readiness, from a high of 89% using a single ANC visit and facility readiness, such as blood pressure apparatus, to a low of less than 1% for all estimates based on receiving eight or more ANC visits. Confidence intervals are wider with the delta method, although for most estimates, both methods provide largely overlapping intervals. At the national level, all intervals were within the bounds of 0 to 1, although for the region-specific estimates (Appendix S6 in the **Online Supplementary Document**) with smaller sample size, the exact method produced 33 (out of 126) invalid confidence intervals that crossed the 0, 1 bounds. These violations affected the estimates near the boundaries: ANC 1 and blood pressure apparatus on the high end, and all estimates based on ANC 8 and/or appropriate hemoglobin and urine protein diagnostic capacity on the low end. The delta method could not be applied in the 24 cases where the adjusted coverage estimate was 0 for a particular type of coverage and a given sub-national region. While the exact method confidence intervals are symmetrical, using the logit transformation in the delta method produces asymmetrical intervals to better reflect the bounds of a proportion. In the case of regions with small sample sizes and adjusted coverage estimates near 1, the lower bound could be quite large, for instance adjusted coverage of 1 ANC visit using blood pressure apparatus as the readiness measure is 97.8%, 95% CI = 15.3%-100%.

Calculating confidence and variance in these adjusted coverage estimates enables inference to be drawn on the results. **Figure 6** shows the delta method results for service readiness based on blood pressure apparatus for four ANC visits. Depicting the variance around the point estimates allows us to identify significant differences between regions and to target the underperforming regions.

In this example, the assumption of independence between regions is likely met as the sampling strategy for both surveys is stratified by region. The assumption of independence between adjusted coverage for facility types within a region is likely not met, because although the SPA facility survey is stratified by facility type (and managing authority), the DHS household survey is not and could not be. However, the covariance between the facility-specific coverage estimates within a region is negative (see Appendix S7 in the **Online Supplementary Document**), which makes the covariance between the facility-specific *effective coverage* estimates within region negative as well. Ignoring this covariance should result in an overestimate of variance and a conser-

Figure 6. Adjusted coverage of antenatal care in 14 regions of Senegal with 95% confidence intervals using the delta method. Results are shown for coverage defined as at least 4 antenatal care visits and readiness defined as functional blood pressure apparatus.

vative confidence interval.

DISCUSSION

Effective coverage is the product of two quantities with their own sampling error, and as such, is an atypical parameter for variance estimation. In this paper, we evaluated the performance of the exact, delta, and bootstrap methods for constructing confidence intervals around effective coverage estimates. The delta method approach performed better than the other methods: it yielded close to nominal coverage, ie, that 95% confidence intervals captured the truth in 95% of the simulations for a given setting, in nearly all cases. The exact and parametric bootstrap methods resulted in undercoverage of nominally 95% confidence intervals – leading to Type I errors in inference – in settings with small sample sizes or coverage and quality measures close to 0 or 1. The exact method also produced invalid confidence intervals in such settings. The non-parametric bootstrap demonstrated consistent undercoverage and was not analyzed in detail.

We applied the delta and exact methods to calculate adjusted coverage of antenatal care in Senegal, using a sample of women with recent live births from the DHS household survey and a sample of health facilities from the SPA facility audit. Both data sources provide estimates with their own sampling error. Using the delta and exact methods enabled us to incorporate both sources of variance. The results demonstrate that the delta and exact methods can be applied to obtain confidence intervals for complex survey data at sub-national and national levels to support inference around levels and trends in adjusted coverage measures. The delta method was more robust in providing valid results even for small sub-national regions; researchers should be aware that confidence intervals may be quite wide for adjusted coverage estimates near 1.

This work shows that it is feasible to characterize the uncertainty around an effective coverage estimate calculated from separate data sources. Variance estimation should become standard practice if effective coverage estimates are to become part of national and global health monitoring [2,5,33]. The methods outlined here may further be applicable to effective coverage estimation in educational settings, such as the human capacity index developed by the World Bank to capture years of quality-adjusted education as a composite measure of years of schooling and achievement test scores [34,35]. Accurate estimation of variability enables stronger inferences that can inform comparisons and targeted improvement actions. Of the methods considered here, we recommend using the delta method approach – because of its superior performance compared to the exact and bootstrap methods, and since the bootstrap approach outlined in this paper has not yet been adapted for effective coverage estimates based on complex survey designs. In order to facilitate use of these methods, we provide code that can be used by researchers for constructing confidence intervals in a variety of settings, including with complex survey data such as the DHS and SPA.

One limitation of this work is that the simulation studies considered the setting in which the quantities used for effective coverage estimation are computed using simple random samples from the data sources; we note that while we expect the relative performance of the methods to hold for quantities estimated using data from a complex sample, the precise coverage levels may differ to a small extent. A second limitation is that to simplify the variance expressions for the exact and delta methods, any covariance between the facility-level or regional-level effective coverage estimates is ignored. In other words, the facility/region-specific effective coverage estimates are taken to be independent in the derivation of the variance formulas. The assumption is not required for cases with a single geographic area and single facility type (or homogeneous quality measures across observed facility types). In the applied example here, where this covariance is nonzero due to dependence of adjusted coverage estimates across different facility types within each region, it is negative and would reduce the variance if accounted for. Investigating the magnitude of this covariance and modifying the methods to take it into account for increased precision are areas for future research.

Researchers should assess the assumptions of independence between region-specific estimates or facility-specific estimates within a region based on the sampling design of the data sources used and note this limitation as relevant. For example, for estimates based on SPA and DHS data, the assumptions for the current analysis would be the same if we further sub-divided facility types by managing authority [30]. However, the assumptions would not hold if adjusted coverage estimates were computed at a geographic level below that used in the sampling schemes for SPA and DHS.

Defining and quantifying variance for effective coverage estimates underscores the atypical elements of this parameter as a product of two proportions or bounded scales. We identified several areas for further investigation on technical elements of the variance. These include modest overcoverage of delta method confidence intervals in the simulation studies with extreme values of coverage and readiness; the assumption that the sampling distribution of F (formula 14, the logit of effective coverage) is approximately normal may not be met in such extreme cases. Further, the poor performance of the non-parametric bootstrap and the lack of parametric bootstrap estimator available for a product of estimates from complex survey data are gaps that warrant development or refinement of bootstrap methods for this setting.

The methods defined here presume that quantities such as facility readiness are measured without sampling error. Further work is required to adapt the confidence interval construction methods to incorporate error in the estimates of coverage or quality. For instance, estimates of quality for adjusted coverage may be based on a sample of health workers [18], observed patient visits [17], or clinical records [36,37]. Researchers must determine if these within-facility measures provide consistent estimates of readiness at the facility level and, if so, how to capture the within-facility sampling error in composite effective coverage.

While incorporating variance in the sampled estimates of coverage and health system quality is important in accurately reporting effective coverage, the methods described here do not incorporate sources of variability that are not captured in the source data. For example, applications of effective coverage estimation frequently combine population data covering a number of years (ANC for live births in the past 2 years would cover visits for pregnancies beginning up to 33 months before the survey) with health system estimates from a single point in time. A particular strength of the data from Senegal used here is that these surveys were conducted simultaneously, but this is more the exception than the rule in existing population and health system surveys. Confidence interval calculation will not address variance introduced from changes in health system measures over time. These sources of uncertainty should be stated in reporting effective coverage results or incorporated via quantitative bias analysis.
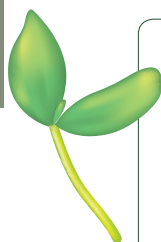
Because measures of crude coverage may obscure the deficits of the health system, assessing effective or adjusted coverage is important in order to recognize both the extent of contact with services, as well as the quality of the services received. Uncertainty estimates that account for the variability in the coverage and quality estimates provide more accurate information for understanding the potential range of quality-adjusted coverage and enabling comparison of these estimates over time and between regions, or countries. Yet, until now, estimates of variance have been largely absent from analysis of effective coverage that combined multiple data sources. This study fills this gap by exploring and comparing three methods for calculating variance. We recommend the delta method as the primary approach for variance calculation, provide Stata code for confidence interval calculation, and outline the assumptions researchers should assess in its application to benefit future research.

REFERENCES

1 United Nations Development Program. Sustainable development goals. Geneva, Switzerland: United Nations, 2015.

2 Kruk ME, Gage A, Arsenault C, Jordan K, Leslie H, Roder-DeWan S, et al. High quality health systems—time for a revolution: Report of the Lancet Global Health Commission on High Quality Health Systems in the SDG Era. Lancet Glob Health. 2018;6:e1196-252. Medline:30196093 doi:10.1016/S2214-109X(18)30386-3

3 Ng M, Fullman N, Dieleman JL, Flaxman AD, Murray CJL, Lim SS. Effective coverage: a metric for monitoring universal health coverage. PLoS Med. 2014;11:e1001730. Medline:25243780 doi:10.1371/journal.pmed.1001730

4 Countdown to 2030 Collaboration. Countdown to 2030: tracking progress towards universal coverage for reproductive, maternal, newborn, and child health. Lancet. 2018;391:1538-48. Medline:29395268 doi:10.1016/S0140-6736(18)30104-1

5 Amouzou A, Leslie HH, Ram M, Fox M, Jiwani S, Requejo J, et al. Advances in the measurement of coverage for RMNCH and nutrition: from Contact to Effective Coverage. BMJ Glob Health. 2019;4 Suppl 4:e001297. Medline:31297252 doi:10.1136/bmjgh-2018-001297

6 Boerma T, AbouZahr C, Evans D, Evans T. Monitoring intervention coverage in the context of universal health coverage. PLoS Med. 2014;11:e1001728. Medline:25243586 doi:10.1371/journal.pmed.1001728

7 Do M, Micah A, Brondi L, Campbell H, Marchant T, Eisele T, et al. Linking household and facility data for better coverage measures in reproductive, maternal, newborn, and child health care: systematic review. J Glob Health. 2016;6:020501. Medline:27606060 doi:10.7189/jogh.06.020501

8 Department of Reproductive Health and Research, Family and Community Health. WHO Antenatal Care Randomized Trial: Manual for the Implementation of the New Model. Geneva, Switzerland: World Health Organization, 2002.

9 United Nations. The Millennium Development Goals Report 2015. New York, New York: United Nations, 2015.

10 World Health Organization. WHO recommendations on antenatal care for a positive pregnancy experience. Geneva, Switzerland: World Health Organization, 2016.

11 Baker U, Okuga M, Waiswa P, Manzi F, Peterson S, Hanson C. Bottlenecks in the implementation of essential screening tests in antenatal care: Syphilis, HIV, and anemia testing in rural Tanzania and Uganda. Int J Gynaecol Obstet. 2015;130:S43-50. Medline:26054252 doi:10.1016/j.ijgo.2015.04.017

12 Baker U, Peterson S, Marchant T, Mbaruku G, Temu S, Manzi F, et al. Identifying implementation bottlenecks for maternal and newborn health interventions in rural districts of the United Republic of Tanzania. Bull World Health Organ. 2015;93:380-9. Medline:26240459 doi:10.2471/BLT.14.141879

13 Heredia-Pi I, Serván-Mori E, Darney BG, Reyes-Morales H, Lozano R. Measuring the adequacy of antenatal health care: a national cross-sectional study in Mexico. Bull World Health Organ. 2016; 94:452-61. Medline:27274597 doi:10.2471/BLT.15.168302

REFERENCES

14 Hodgins S, D'Agostino A. The quality-coverage gap in antenatal care: toward better measurement of effective coverage. Glob Health Sci Pract. 2014;2:173-81. Medline:25276575 doi:10.9745/GHSP-D-13-00176

15 Kanyangarara M, Munos MK, Walker N. Quality of antenatal care service provision in health facilities across sub–Saharan Africa: Evidence from nationally representative health facility assessments. J Glob Health. 2017;7:021101. Medline:29163936 doi:10.7189/jogh.07.021101

16 Kiwanuka Henriksson D, Fredriksson M, Waiswa P, Selling K, Swartling Peterson S. Bottleneck analysis at district level to illustrate gaps within the district health system in Uganda. Glob Health Action. 2017;10:1327256. Medline:28581379 doi:10.1080/16549716.2017.1327256

17 Leslie HH, Malata A, Ndiaye Y, Kruk ME. Effective coverage of primary care services in eight high-mortality countries. BMJ Glob Health. 2017;2:e000424. Medline:29632704 doi:10.1136/bmjgh-2017-000424

18 Marchant T, Tilley-Gyado RD, Tessema T, Singh K, Gautham M, Umar N, et al. Adding content to contacts: Measurement of high quality contacts for maternal and newborn health in Ethiopia, North East Nigeria, and Uttar Pradesh, India. PLoS One. 2015;10:e0126840. Medline:26000829 doi:10.1371/journal.pone.0126840

19 Nguhiu PK, Barasa EW, Chuma J. Determining the effective coverage of maternal and child health services in Kenya, using demographic and health survey data sets: tracking progress towards universal health coverage. Trop Med Int Health. 2017;22:442-53. Medline:28094465 doi:10.1111/tmi.12841

20 World Health Organization. World Bank. Tracking Universal Health Coverage: First Global Monitoring Report. World Health Organization, World Bank, 2015.

21 Colson KE, Zúñiga-Brenes P, Ríos-Zertuche D, Conde-Glez CJ, Gagnier MC, Palmisano E, et al. Comparative estimates of crude and effective coverage of measles immunization in low-resource settings: Findings from Salud Mesoamérica 2015. PLoS One. 2015;10:e0130697. Medline:26136239 doi:10.1371/journal.pone.0130697

22 Aaron GJ, Strutt N, Boateng NA, Guevarra E, Siling K, Norris A, et al. Assessing program coverage of two approaches to distributing a complementary feeding supplement to infants and young children in Ghana. PLoS One. 2016;11:e0162462. Medline:27755554 doi:10.1371/journal.pone.0162462

23 Wang W, Mallick L, Allen C, Pullam T. Effective coverage of facility delivery in Bangladesh, Haiti, Malawi, Nepal, Senegal, and Tanzania. PLoS One. 2019;14:e0217853. Medline:31185020 doi:10.1371/journal.pone.0217853

24 Willey B, Waiswa P, Kajjo D, Munos M, Akuze J, Allen E, et al. Linking data sources for measurement of effective coverage in maternal and newborn health: what do we learn from individual- vs ecological-linking methods? J Glob Health. 2018;8:010601. Medline:29497508 doi:10.7189/jogh.06.0207028.010601

25 Munos MK, Maiga A, Do M, Sika GL, Carter ED, Mosso R, et al. Linking household survey and health facility data for effective coverage measures: a comparison of ecological and individual linking methods using the Multiple Indicator Cluster Survey in Côte d'Ivoire. J Glob Health. 2018;8:020803. Medline:30410743 doi:10.7189/jogh.08.020803

26 Wang W, Mallick L, Allen C, Pullum T. Effective coverage of facility delivery in Bangladesh, Haiti, Malawi, Nepal, Senegal, and Tanzania. Rockville, Maryland: ICF; 2018.

27 Goodman LA. On the Exact Variance of Products. J Am Stat Assoc. 1960;55:708-13. doi:10.1080/01621459.1960.10483369

28 Hogg RV, Craig AT. Introduction to mathematical statistics. 2nd ed. New York: Macmillan Publishing Co., Inc.; 1965.

29 Efron B. Bootstrap methods: Another look at the jackknife. Ann Stat. 1979;7:1-26. doi:10.1214/aos/1176344552

30 Agence Nationale de la Statistique et de la Démographie (ANSD) [Sénégal], ICF International. Sénégal: Enquête Continue sur la Prestation des Services de Soins de Santé (ECPSS) 2014. Rockville, Maryland: Agence Nationale de la Statistique et de la Démographie (ANSD) [Sénégal] and ICF International, 2015.

31 World Health Organization. Service Availability and Readiness Assessment (SARA) Implementation Guide. Geneva, Switzerland: World Health Organization, 2015 September 2015. Report No.: Contract No.: Version 2.2.

32 Lahiri P. On the impact of bootstrap in survey sampling and small-area estimation. Stat Sci. 2003;18:199-210. doi:10.1214/ss/1063994975

33 Lozano R, Soliz P, Gakidou E, Abbott-Klafter J, Feehan DM, Vidal C, et al. Benchmarking of performance of Mexican states with effective coverage. Lancet. 2006;368:1729-41. Medline:17098091 doi:10.1016/S0140-6736(06)69566-4

34 World Bank. World Development Report 2019: The Changing Nature of Work. Washington DC: The World Bank, 2019.

35 World Bank. World Development Report 2018: Learning to Realize Education's Promise. Washington DC: The World Bank, 2018.

36 Chaturvedi S, Upadhyay S, De Costa A, Raven J. Implementation of the partograph in India's JSY cash transfer programme for facility births: a mixed methods study in Madhya Pradesh province. BMJ Open. 2015;5:e006211. Medline:25922094 doi:10.1136/bmjopen-2014-006211

37 Khan ANS, Billah SM, Mannan I, Mannan II, Begum T, Khan MA, et al. A cross-sectional study of partograph utilization as a decision making tool for referral of abnormal labour in primary health care facilities of Bangladesh. PLoS One. 2018;13:e0203617. Medline:30188940 doi:10.1371/journal.pone.0203617