Estimation of sensitivity and specificity and calculation of sample size for a validation study with stratified sampling

Kiyoshi Kubota (kubotape-tky@umin.net)
NPO Drug Safety Research Unit Japan https://orcid.org/0000-0001-6092-8685
Masao Iwagami
Tsukuba Daigaku
Takuhiro Yamaguchi
Tohoku Daigaku

Research article

Keywords: Sensitivity, Specificity, Predictive values, Sample size

Posted Date: August 13th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-55913/v1

Research Article

Estimation of sensitivity and specificity and calculation of sample size for a validation study with stratified sampling

Kiyoshi Kubota[1] Masao Iwagami[2,3] and Takuhiro Yamaguchi[4]

[1]NPO Drug Safety Research Unit Japan, Tokyo, Japan, kubotape-tky@umin.net

[2]Department of Health Services Research, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan, iwagami-tky@umin.ac.jp

[3]Department of Non-Communicable Disease Epidemiology, London School of Hygiene, Tropical Medicine, London, United Kingdom

[4]Division of Biostatistics, Tohoku University Graduate School of Medicine, Miyagi, Japan, yamaguchi@med.tohoku.ac.jp

Correspondence:

Kiyoshi Kubota, MD PhD FISPE,

NPO Drug Safety Research Unit Japan, Yushima 1-2-13-4F, Bunkyo-ku, Tokyo 113-0034 Japan.

Tel: +81-3-5297-5860

Fax: +81-3-5297-5890

E-mail: kubotape-tky@umin.net

Abstract

Background:

We propose and evaluate the approximation formulae for the 95% confidence intervals (CIs) of the sensitivity and specificity and a formula to estimate sample size in a validation study with stratified sampling where positive samples satisfying the outcome definition and negative samples that do not are selected with different extraction fractions.

Methods:

We used the delta method to derive the approximation formulae for estimating the sensitivity and specificity and their CIs. From those formulae, we derived the formula to estimate the size of negative samples required to achieve the intended precision and the formula to estimate the precision for a negative sample size arbitrarily selected by the investigator. We conducted simulation studies in a population where 4% were outcome definition positive, the positive predictive value (PPV)=0.8, and the negative predictive value (NPV)=0.96, 0.98 and 0.99. The size of negative samples, n_0, was either selected to make the 95% CI fall within ± 0.1, 0.15 and 0.2 or set arbitrarily as 150, 300 and 600. We assumed a binomial distribution for the positive and negative samples. The coverage of the 95% CIs of the sensitivity and specificity was calculated as the proportion of CIs including the sensitivity and specificity in the population, respectively. For selected studies, the coverage was also estimated by the bootstrap method. The sample size was evaluated by examining whether the observed precision was within the pre-specified value.

Results:

For the sensitivity, the coverage of the approximated 95% CIs was larger than 0.95 in most studies but in 9 of 18 selected studies derived by the bootstrap method. For the specificity, the coverage of the approximated 95% CIs was approximately 0.93 in most studies, but the coverage was more than 0.95 in all 18 studies derived by the bootstrap method. The calculated size of negative samples yielded precisions within the pre-specified values in most of the studies.

Conclusion:

The approximation formulae for the 95% CIs of the sensitivity and specificity for stratified validation studies are presented. These formulae will help in conducting and analysing validation studies with stratified sampling.

Keywords: Sensitivity; Specificity; Predictive values; Sample size

Background

In studies constructed to evaluate the validity of an outcome definition implementing chart reviews as the gold standard, the chart review is the most time-consuming part of the study. This is one of the main reasons why many validation studies estimate the positive predictive value (PPV) only when using chart reviews as the gold standard [1]. For example, in a recent review of 14 validation studies on dementia diagnoses, McGuinness stated, “Most reported only the positive predictive value (PPV)” [2]. In a paper published in the FDA-funded Mini-Sentinel pilot programme, the authors stated, “It was determined that 100 charts would be sufficient to obtain a reasonable PPV and establish the validation process” [3]. There are, however, several validation studies where many charts are reviewed to estimate not only the positive predictive value (PPV) but also other measures of validity (negative predictive value (NPV), sensitivity and specificity). For example, Widdifield et al. used as many as 9,500 random samples to identify 107 patients with rheumatoid arthritis (RA): 7,500 random samples of patients aged ≥20 years to identify 69 RA patients and an additional 2,000 aged ≥65 years to identify 38 RA patients [4]. Instead of employing the random sampling strategy, one may use the stratified sampling strategy, where positive samples obtained among patients who satisfy the outcome definition and negative samples obtained from patients who do not satisfy the definition are selected with different extraction fractions. In validation studies using the stratified sampling strategy, however, there seems to be no generally accepted approach to estimate measures of validity. For example, in a validation study published in 2014 by Husain et al. on nonalcoholic fatty liver disease, the sensitivity and specificity were estimated in a "population" of 600 patients, consisting of 450 who satisfied the outcome definition and 150 who did not [5]. One reason why the measures of validity were estimated in this artificial "population" but not in the original population may be that there was no generally accepted approach to estimate the sensitivity and specificity and their confidence intervals (CIs) in the original population when the stratified sampling strategy was employed.

One valid approach to obtaining these estimates is the use of the bootstrap method to estimate the CIs of the sensitivity and specificity as in a validation study of preeclampsia in Norway published in 2014 by Klungsøyr et al. [6]. In this study, all (i.e., 100%) of the 3,500 women in a pregnancy cohort registered with preeclampsia (positive samples) and 1,840 (2.4%) random samples from 75,311 women without registered preeclampsia (negative samples) were examined utilizing antenatal charts and hospital discharge codes used as the gold standard. The CIs of the sensitivity and specificity in the original population of 78,811 women were estimated by the bootstrap method rather than the CIs in the artificial

"population" of 5,340 women [6].

In the current study, we present approximation formulae for the CIs of the sensitivity and specificity in the original population in a validation study with stratified sampling. The formulae may be useful when the standard statistical package for the bootstrap method is not readily available. We also propose a relevant formula to estimate the size of negative samples required to achieve the intended precision or to estimate the size of the precision when an arbitrary negative sample size is used, provided that the PPV has already been estimated from positive samples.

Methods

Approximated 95% confidence intervals (CIs) of the sensitivity and specificity from a validation study with stratified sampling

The approximated 95% CIs of sensitivity and specificity were derived using the formula for the logarithm of the risk ratio [7] and the delta method (see Additional file 1 for the derivation) as follows.

The 95% CI of the sensitivity, $95\%CI_{se}$, is approximated as:

$$95\%CI_{se} = \widehat{se} \cdot exp\left[\pm 1.96(1-\widehat{se})\sqrt{\frac{1}{a}-\frac{1}{n_1}+\frac{1}{c}-\frac{1}{n_0}}\right] \quad (1)$$

Similarly, the 95% CI of the specificity, $95\%CI_{sp}$, is approximated as:

$$95\%CI_{sp} = \widehat{sp} \cdot exp\left[\pm 1.96(1-\widehat{sp})\sqrt{\frac{1}{b}-\frac{1}{n_1}+\frac{1}{d}-\frac{1}{n_0}}\right] \quad (2)$$

In Equations (1) and (2), 'a' is the number of true positives (TPs) and 'b' is the number of false positives (FPs) among positive samples of n_1 subjects (n_1=a+b), while 'c' is the number of false negatives (FNs) and 'd' is the number of true negatives (TNs) among negative samples of n_0 subjects (n_0=c+d). The expected point estimate of the sensitivity ($\widehat{se}$) in Equation (1) is approximated as:

$$\widehat{se} \sim \frac{N_1 a/n_1}{N_1 a/n_1 + N_0 c/n_0} = \frac{\widehat{PPV}}{\widehat{PPV}+(1-\widehat{NPV})N_0/N_1} \quad (3)$$

Similarly, the expected point estimate of the specificity ($\widehat{sp}$) in Equation (2) is approximated

as:

$$\widehat{sp} \sim \frac{N_0 d/n_0}{N_1 b/n_1 + N_0 d/n_0} = \frac{\widehat{NPV} N_0/N_1}{(1-\widehat{PPV}) + \widehat{NPV}\, N_0/N_1} \quad (4)$$

In Equations (3) and (4), $\widehat{PPV}$ is the estimate of the positive predictive value in positive samples, given as $\widehat{PPV}$= a/n_1, $\widehat{NPV}$ is the negative predictive value in negative samples, given as $\widehat{NPV}$= d/n_0=(1- c/n_0), N_1 is the number of subjects who satisfy the outcome definition, and N_0 is the number of subjects who do not in the original population.
Alternative formulae for $95\%CI_{se}$ and $95\%CI_{sp}$ worth exploring are given as:
:

$95\%CI_{se} = \widehat{se} \pm \delta_{se}$

where δ_{se} is given as

$$\delta_{se} = \widehat{se} \cdot \left(exp\left[1.96(1-\widehat{se}) \sqrt{\frac{1}{a} - \frac{1}{n_1} + \frac{1}{c} - \frac{1}{n_0}} \right] - 1 \right) \quad (5)$$

and $95\%CI_{sp} = \widehat{sp} \pm \delta_{sp}$

where δ_{sp} is given as

$$\delta_{sp} = \widehat{sp} \cdot \left(exp\left[1.96(1-\widehat{sp}) \sqrt{\frac{1}{b} - \frac{1}{n_1} + \frac{1}{d} - \frac{1}{n_0}} \right] - 1 \right) \quad (6)$$

When comparing the calculation of $95\%CI_{se}$ with Equations (1) and (5), the upper limit is the same, but the lower limit in Equation (5) is lower than that in Equation (1). Similarly, the upper limit of $95\%CI_{sp}$ in Equation (2) is the same as that in Equation (6), but the lower limit in Equation (6) is lower than that in Equation (2).

Estimation of the negative sample size and precision in a validation study with stratified sampling

We assume a validation study where the primary interest is to estimate the PPV of the outcome definition. When the PPV estimated by chart reviews of positive samples ($\widehat{PPV}$) is promising (e.g., $\widehat{PPV}$ is higher than 0.7 or 0.8 [8-10]), it is hoped that some information is obtained for other measures of validity, particularly for the sensitivity but also for the NPV and specificity, even if the precision for some of those measures may be lower than that for the PPV. In the current article, we define the first stage, in which the PPV is estimated from

positive samples, as 'Stage I' and the next stage, in which the NPV, sensitivity and specificity are estimated, as 'Stage II'. In the proposed method, the chart reviews in Stage I should be completed and 'a' and $\widehat{PPV}$ should be estimated before the negative samples are selected in Stage II. The size of negative samples (n_0) in Stage II is calculated (see Additional file 1 for the derivation) as follows:

$$n_0 = \left(\frac{N_0}{N_1}\frac{1}{\widehat{PPV}}\frac{se^*}{1-se^*} - 1\right)\left[\left(\frac{\log\left(1+\frac{\delta_{se}{}^*}{se^*}\right)}{1.96(1-se^*)}\right)^2 - \frac{1}{a}(1-\widehat{PPV})\right]^{-1} \quad (7)$$

In Equation (7), N_0/N_1 is obtained from the information on the population, the values of 'a' and $\widehat{PPV}$ are obtained in Stage I, $\delta_{se}{}^*$ is the precision of $\widehat{se}$ that must be obtained (i.e., the $95\%CI_{se}$ obtained at the end of the study will be $\widehat{se} \pm \delta_{se}{}^*$ or narrower) and se^* is the sensitivity used to calculate the size of negative samples (n_0). If good information on the sensitivity (e.g., the sensitivity estimated in a past study conducted in a similar population) is available, a likely value of the sensitivity may be used as se^* in Equation (7). However, if no good information on the sensitivity is available, one may use se^*, which produces the possible largest value of n_0 (defined as n_{0max}). The value of n_{0max} can be determined numerically. Alternatively, se^* in Equation (7) may be fixed as 0.7 because n_0 in Equation (7) varies with se^* but is maximal (or n_{0max}) when se^* is approximately 0.7 (see Additional File 1). The approximated largest value of n_0, defined as $n_0(0.7)$, is given as:

$$n_0(0.7) = \frac{2.33\frac{N_0}{N_1}\frac{1}{\widehat{PPV}} - 1}{\left[2.89\left(\log\left(1+\frac{\delta_{se}{}^*}{0.7}\right)\right)^2 - \frac{1}{a}(1-\widehat{PPV})\right]} \quad (8)$$

Conversely, when the size of negative samples, n_0, is arbitrarily specified by the investigator (defined as $n_0{}^*$), the precision δ_{se} can be predicted as (see Additional file 1 for the derivation):

$$\delta_{se} = se^{**}\left(exp\left[1.96(1-se^{**})\sqrt{\frac{1}{a}(1-\widehat{PPV}) + \left(\frac{se^{**}}{1-se^{**}}\frac{N_0}{N_1}\frac{1}{\widehat{PPV}} - 1\right)\frac{1}{n_0{}^*}}\right] - 1\right) \quad (9)$$

In Equation (9), as in Equations (7) and (8), N_0/N_1 is obtained from the information on the population, $\widehat{PPV}$ and 'a' are obtained in Stage I and se^{**} is the sensitivity used to estimate δ_{se}. In Equation (9), the value of $\delta_{se}$is maximized when se^{**} is approximately 0.7 (defined

as δ_{se_max}) and $\delta_{se}(0.7)$, the value of δ_{se} obtained when se^{**} is fixed as 0.7, can be used as an estimate of δ_{se_max} (see additional File 1). The value of $\delta_{se}(0.7)$ is given as:

$$\delta_{se}(0.7) = 0.7\left(exp\left[0.588\sqrt{\frac{1}{a}(1-\widehat{PPV}) + (2.33\frac{N_0}{N_1}\frac{1}{\widehat{PPV}} - 1)\frac{1}{n_0{}^*}}\right] - 1\right) \qquad (10)$$

Numerical examples

To evaluate the formulae proposed in the preceding two sections, 36 fictitious validation studies were examined (Tables 1 and 2). We assumed that in the original population, N_1 outcome definition-positive subjects accounted for 4% and N_0 outcome definition-negative subjects accounted for the remaining 96% of the population (N_0/N_1=24). Among the N_1 definition-positive subjects, 80% were TPs and 20% were FPs (the positive predictive value in the population (PPV) is 0.8). We considered 3 different values (0.99, 0.98 and 0.96) for the proportion of TNs (the negative predictive value (NPV)) in N_0 definition-negative subjects in the population). When SE and SP are defined as the sensitivity and specificity in the population, SE=0.769 and SP=0.992 when NPV=0.99 (Studies 1-9 in Table 1 and Studies 28-30 in Table 2), SE=0.625 and SP=0.992 when NPV=0.98 (Studies 10-18 and Studies 31-33), and SE=0.455 and SP=0.991 when NPV=0.96 (Studies 19-27 and Studies 31-33).
We simulated two-stage validation studies with stratified sampling. In Stage I of all studies, 100 (n_1) random positive samples were selected from N_1 definition-positive subjects and evaluated by chart reviews. After the completion of chart reviews, $\widehat{PPV}$ was estimated as a/n_1. Then, the size of negative samples (n_0) in Stage II was determined by one of the following 4 options. In Option A, n_{0max}, or the largest value of n_0 that can achieve the pre-specified precision ($\delta_{se}{}^*$), is calculated by Equation (7); in Option B, $n_0(0.7)$ (where se^* is fixed as 0.7) is obtained by Equation (8); in Option C, n_0 is calculated by Equation (7), where the information on the population sensitivity (SE) available from the previous study is used as se^*; and in Option D, n_0 is arbitrarily selected by the investigator. For Options A-C, the size of negative samples (n_0) was selected to attain $\delta_{se}{}^*$=0.2, 0.15 and 0.1 in Equation (7) or (8). In Option C, it was assumed that based on the information from the previous study, se^* in Equation (7) was set as 0.8 in Studies 3, 6 and 9, where SE=0.769, se^* was set as 0.6 in Studies 12, 15 and 18, where SE=0.625 and se^* was set as 0.5 in Studies 21, 24 and 27, where SE=0.455. In Option D, it was assumed that the investigator had selected 150, 300 or 600 random negative samples, or $n_0{}^*$=150, 300 or 600, in Equation (9) or (10).

In all the studies, the number of TPs (a) among 100 (n_1) positive samples and the number of FNs (c) among n_0 negative samples were assumed to follow a binomial distribution. In Stage I, the distribution of 'a' was a ~ B(100, 0.8), and in Stage II, the distribution of 'c' was c

~ B(n_0, 0.01) in Studies 1-9 and 28-30, c ~ B(n_0, 0.02) in Studies 10-18 and 31-33 and c ~ B(n_0, 0.04) in Studies 19-27 and 34-36.

We used the RAND function of SAS 9.4 with 10,000 iterations for each of the 36 studies. The sensitivity and specificity were estimated by Equations (3) and (4), and their 95% CIs were estimated by Equations (1) and (2). We also estimated the alternative 95% CIs in Equations (5) and (6). We estimated the coverage of $95\%CI_{se}$ as the proportion of 10,000 iterations in which SE was within $95\%CI_{se}$ in Equation (1) or (5). Similarly, we estimated the coverage of $95\%CI_{sp}$ as the proportion of 10,000 iterations in which SP was included within $95\%CI_{sp}$ in Equation (2) or (6). When the upper limit of $95\%CI_{se}$ in Equation (1) or (5) exceeded 1, the upper limit was set to 1. When c=0, both $\widehat{se}$ and the upper limit of $95\%CI_{se}$ were set to 1, and the lower limit of $95\%CI_{se}$ was calculated as $PPV/(PPV + (3/n_0)(N_0/N_1))$. The formula $PPV/(PPV + (3/n_0)(N_0/N_1))$ was derived from Equation (3), where 3/n_0 was used as c/n_0 $(= 1 - \widehat{NPV})$ according to the "rule of three" [11].

We also examined the coverage of $95\%CI_{se}$ and $95\%CI_{sp}$ by using the bootstrap method for Studies 1, 4, 7, 10, 13, 16, 19, 22, and 25 (Option A) and Studies 28, 29, --, and 33 (Option D). In the bootstrap method, we selected 1,000 resamples with SAS 9.4 SURVEYSELECT for each of 10,000 iterations. For the bootstrap method, the coverage was estimated as the proportion of 1,000 resamples in which SE was included in the 2.5th-97.5th percentile range of $\widehat{se}$ estimated by equation (3) and as the proportion of 1,000 resamples in which SP was included in the 2.5th-97.5th percentile range of $\widehat{sp}$ estimated by equation (4).

As the coverage of $95\%CI_{se}$ or $95\%CI_{sp}$ in Equations (1), (2), (5) and (6) was often found to be less than 0.95 (see Results), we conducted post hoc studies to simulate a total of 10,560 fictitious studies with different combinations of ${\delta_{se}}^*$, PPV, NPV, N_0/N_1 and n_1 and calculated n_0 in Equation (7) with Option A and examined the coverage of $95\%CI_{se}$ in Equations (1) and (5) and the coverage of $95\%CI_{sp}$ in Equations (2) and (6). We used {0.1, 0.15, 0.2, 0.25} for ${\delta_{se}}^*$, {0.5, 0.6, 0.7, 0.8, 0.9, 0.95} for PPV, {0.8, 0.9, 0.93, 0.96, 0.98, 0.99, 0.995, 0.9975} for NPV, {9, 19, 24, 49, 99} for N_0/N_1 and {60, 80, 100, 120, 150, 200, 300, 500, 800, 1000, 2000} for n_1.

Last, we compared $95\%CI_{se}$ calculated from Equations (1) and (5) as well as $95\%CI_{sp}$ calculated from Equations (2) and (6) with the bootstrap percentiles shown in the study by Klungsøyr et al. [6].

The current study investigates formulae and designs in validation studies where only simulation data are used. Therefore, no ethics review was needed for the study.

Results

Table 1 shows the medians and 2.5th-97.5th percentile ranges of $\widehat{se}$ and $\widehat{sp}$ obtained over 10,000 iterations. The medians of $\widehat{se}$ and $\widehat{sp}$ are close to SE and SP, respectively. Table 1 also indicates that a larger negative sample size (n_0) is required to attain higher precision (smaller ${\delta_{se}}^*$), and additionally, the largest possible n_0 from Option A is close to $n_0(0.7)$ from Option B. Table 2 shows that as the size of negative samples (${n_0}^*$)used in Option D increases, the predicted precision (δ_{se_max} or $\delta_{se}(0.7)$) decreases. It also shows that δ_{se_max} is close to $\delta_{se}(0.7)$.

Table 3 shows the medians of $\widehat{se}$, the lower and upper limits of $95\%CI_{se}$, and the values of δ_{se} for 18 studies (Options A and D). The median δ_{se} (the value calculated in Equation (5) for $95\%CI_{se}$ or 1- $\widehat{se}$ when the upper limit is set to 1 in Equations (1) and (5), and the difference between the 97.5th percentile and the median $\widehat{se}$ among 1,000 resamples in the bootstrap method) is similar to but smaller than the intended size of the precision (${\delta_{se}}^*$) from Option A (Studies 1, 4, ---, 25). Likewise, the median of δ_{se} is smaller than δ_{se_max} from Option D (Studies 28, 29, ---, 36), particularly in Studies 28, 31 and 34, where ${n_0}^*$=150. The coverage of $95\%CI_{se}$ calculated in Equation (1) is less than 0.95 in 15 of 18 studies, while the coverage of $95\%CI_{se}$ calculated in Equation (5) is less than 0.95 in 3 of 18 studies. The coverage of the 2.5th-97.5th percentile range of $\widehat{se}$ estimated by the bootstrap method is less than 0.95 in 9 of 18 studies. In one study where ${n_0}^*$=150 (Study 28), the coverage of the bootstrap method is as low as 0.778. Table 4 shows the medians of $\widehat{sp}$, the lower and upper limits of $95\%CI_{sp}$ and δ_{sp} (the value calculated in Equation (6) for $95\%CI_{sp}$ calculated in Equations (2) and (6) or the difference between the 97.5th percentile and median $\widehat{sp}$ among 1,000 resamples calculated in the bootstrap method) for 18 studies. The coverage of $95\%CI_{sp}$ estimated by Equations (2) and (6) is less than 0.95 (approximately 0.93) in all 18 studies, and there is no substantial difference in the coverage of $95\%CI_{sp}$ between Equations (2) and (6). On the other hand, the coverage of the 2.5th-97.5th percentile range of $\widehat{sp}$ estimated by the bootstrap method is larger than 0.95 in all 18 studies. The 2.5th-97.5th percentile range in the bootstrap method of $\widehat{sp}$ is wider than $95\%CI_{sp}$, and the precision of $\widehat{sp}$ (difference between the 97.5 percentile and the median) in the bootstrap method is larger than that of δ_{sp} from Equation (6).

Table 5 shows the coverage of $95\%CI_{se}$ and $95\%CI_{sp}$ from 10,560 post hoc studies. The coverage of $95\%CI_{se}$ is larger than 0.95 in 38.5% of the 10,560 studies when Equation (1) is used but is larger than 0.95 in 84.2% when Equation (5) is used. On the other hand, the coverage of $95\%CI_{sp}$ is larger than 0.95 only in 2.5 and 3.4% of the studies, respectively,

when Equation (2) and Equation (6) are used. However, $95\%CI_{sp}$ from Equation (2) or (6) is larger than 0.93 in more than 80% of the 10,560 studies.

Last, for the data presented by Klungsøyr et al. [6], Equations (1) and (2) gave a sensitivity of 43.0 (95% CI: 38.5, 48.1)% and a specificity of 99.2 (99.2, 99.3)%, while Equations (5) and (6) gave a sensitivity of 43.0 (37.9, 48.1)% and a specificity of 99.2 (99.2, 99.3)%. Those estimates are close to the bootstrap estimates, 43.0 (38.7, 48.2)% for the sensitivity and 99.2 (99.2, 99.3)% for the specificity, presented in study [6].

Discussion

When stratified sampling is employed in a validation study, the measures of validity in the original population rather than those in the artificial "population" should be estimated and reported to provide useful information. This is because the validated definition is normally used in the study conducted with the original population. For example, in the study by Klungsøyr et al [6], the sensitivity was 96.8%, and the specificity was 75.6% for the artificial "population" of 5,340 women, which was different from the sensitivity (43.0%) and specificity (99.3%) estimated for the original population.

Our study indicated that the 95% CI of the sensitivity and specificity can be approximated by Equations (1) and (2), respectively. However, we recommend the use of Equation (5) rather than Equation (1) to estimate the 95% CI for the sensitivity because the coverage of $95\%CI_{se}$ from Equation (1) is larger than 0.95 in less than half of the studies shown in Table 3 and Table 5, while the coverage is larger than 0.95 in 15 of 18 studies shown in Table 3 and in more than 90% shown in Table 5. The coverage of $95\%CI_{sp}$ was less than 0.95 when Equation (2) or (6) was used. Table 4 implies that the bootstrap method may give a better estimate for the CI of the specificity. However, Equation (2) or (6) may still be of use because the coverage was at least 0.93 in more than 80% of the studies shown in Table 5.

We also presented a formula to calculate the size of negative samples required to attain an intended precision once the chart reviews in Stage I are completed before selecting negative samples in Stage II. In addition, we presented a formula to calculate the precision that would be attained when the investigator uses a predetermined negative sample size.

The approximation formulae for $95\%CI_{se}$ and $95\%CI_{sp}$ and formulae to calculate the size of negative samples and attainable precision may encourage the conduction of validation studies with stratified sampling, which can provide some information on sensitivity, NPV and specificity even if the primary purpose of the study is to estimate the PPV of the outcome definition.

Conclusion

We proposed formulae to approximate the 95% CIs of the sensitivity and specificity for validation studies with stratified sampling. We also proposed a formula to estimate the size of negative samples required to attain a pre-specified precision. The formulae may help in the proper conduction and analysis of validation studies with stratified sampling.

Table 1 Validation studies with stratified sampling where the precision of the sensitivity ($\delta_{se}{}^{*}$) is specified: Monte Carlo simulations with 10,000 iterations

Study ID	Option	$\delta_{se}{}^{*}$	se^{*}	n_0	$\widehat{PPV}$	$\widehat{NPV}$	$\widehat{se}$	$\widehat{sp}$
N_0/N_1=24, PPV=0.8, NPV=0.99, SE=0.769, SP=0.992, n_1=100 (Study ID= 1-6)								
1	A	0.2	se_0	388 (350-434)[a]	0.800 (0.720-0.880)[a]	0.990 (0.979-0.998)[a]	0.764 (0.617-0.929)[a]	0.992 (0.988-0.995)[a]
2	B	0.2	0.7	383 (346-430)	0.800 (0.720-0.880)	0.990 (0.979-0.998)	0.762 (0.615-0.928)	0.992 (0.988-0.995)
3	C	0.2	0.8	371 (340-414)	0.800 (0.720-0.870)	0.990 (0.978-0.998)	0.756 (0.607-0.925)	0.992 (0.988-0.995)
4	A	0.15	se_0	656 (591-738)	0.800 (0.720-0.880)	0.990 (0.982-0.997)	0.760 (0.648-0.917)	0.992 (0.988-0.995)
5	B	0.15	0.7	648 (582-731)	0.800 (0.720-0.880)	0.990 (0.982-0.997)	0.781 (0.660-0.916)	0.992 (0.988-0.995)
6	C	0.15	0.8	628 (574-703)	0.800 (0.720-0.870)	0.990 (0.982-0.998)	0.777 (0.655-0.913)	0.992 (0.988-0.995)
7	A	0.1	se_0	1422 (1267-1622)	0.800 (0.720-0.880)	0.990 (0.985-0.995)	0.771 (0.686-0.872)	0.992 (0.988-0.995)
8	B	0.1	0.7	1407 (1248-1611)	0.800 (0.720-0.880)	0.990 (0.985-0.995)	0.769 (0.687-0.871)	0.992 (0.988-0.995)
9	C	0.1	0.8	1356 (1232-1532)	0.800 (0.720-0.870)	0.990 (0.985-0.995)	0.765 (0.683-0.882)	0.992 (0.988-0.995)
N_0/N_1=24, PPV=0.8, NPV=0.98, SE=0.625, SP=0.992, n_1=100 (Study ID= 7-12)								
10	A	0.2	se_0	388 (350-434)	0.800 (0.720-0.880)	0.980 (0.966-0.994)	0.618 (0.498-0.865)	0.992 (0.988-0.995)
11	B	0.2	0.7	383 (346-430)	0.800 (0.720-0.880)	0.980 (0.966-0.994)	0.616 (0.495-0.864)	0.992 (0.988-0.995)
12	C	0.2	0.6	333 (300-375)	0.800 (0.720-0.880)	0.980 (0.965-0.994)	0.615 (0.482-0.848)	0.992 (0.988-0.995)
13	A	0.15	se_0	656 (591-738)	0.800 (0.720-0.880)	0.980 (0.969-0.991)	0.627 (0.522-0.784)	0.992 (0.988-0.995)
14	B	0.15	0.7	648 (582-731)	0.800 (0.720-0.880)	0.980 (0.969-0.991)	0.624 (0.520-0.782)	0.992 (0.988-0.995)
15	C	0.15	0.6	561 (508-636)	0.800 (0.720-0.870)	0.980 (0.968-0.991)	0.629 (0.511-0.790)	0.992 (0.988-0.995)

Table 1 --- continued

16	A	0.1	se_0	1422 (1267-1622)	0.800 (0.720-0.880)	0.980 (0.973-0.987)	0.625 (0.547-0.725)	0.992 (0.988-0.995)
17	B	0.1	0.7	1407 (1248-1611)	0.800 (0.720-0.880)	0.980 (0.973-0.987)	0.625 (0.547-0.724)	0.992 (0.988-0.995)
18	C	0.1	0.6	1217 (1087-1410)	0.800 (0.720-0.870)	0.980 (0.972-0.988)	0.625 (0.543-0.739)	0.992 (0.988-0.995)
N_0/N_1=24, PPV=0.8, NPV=0.96, SE=0.455, SP=0.991, n_1=100 (Study ID= 19-27)								
19	A	0.2	se_0	388 (350-434)	0.800 (0.720-0.880)	0.960 (0.940-0.979)	0.449 (0.360-0.618)	0.991 (0.988-0.995)
20	B	0.2	0.7	383 (346-430)	0.800 (0.720-0.880)	0.960 (0.940-0.979)	0.459 (0.357-0.615)	0.991 (0.988-0.995)
21	C	0.2	0.5	252 (229-284)	0.800 (0.720-0.870)	0.961 (0.934-0.983)	0.457 (0.332-0.675)	0.991 (0.988-0.995)
22	A	0.15	se_0	656 (591-738)	0.800 (0.720-0.880)	0.960 (0.945-0.975)	0.456 (0.374-0.576)	0.991 (0.988-0.995)
23	B	0.15	0.7	648 (582-731)	0.800 (0.720-0.880)	0.960 (0.945-0.975)	0.454 (0.375-0.574)	0.991 (0.988-0.995)
24	C	0.15	0.5	420 (374-478)	0.800 (0.720-0.880)	0.960 (0.941-0.979)	0.453 (0.359-0.607)	0.991 (0.988-0.995)
25	A	0.1	se_0	1422 (1267-1622)	0.800 (0.720-0.880)	0.960 (0.950-0.970)	0.454 (0.395-0.530)	0.991 (0.988-0.995)
26	B	0.1	0.7	1407 (1248-1611)	0.800 (0.720-0.880)	0.960 (0.950-0.970)	0.455 (0.395-0.531)	0.991 (0.988-0.995)
27	C	0.1	0.5	904 (803-1053)	0.800 (0.720-0.870)	0.960 (0.948-0.973)	0.454 (0.383-0.554)	0.991 (0.988-0.994)

a Medians and 2.5th-97.5th percentile range (in parentheses).

Option: option to select n_0 (see text for the details); $\delta_{se}{}^*$: intended precision of the sensitivity; se^*: value used in Equation (7); n_0: size of negative samples; $\widehat{PPV}$: positive predictive value in the samples; $\widehat{NPV}$: negative predictive value in the samples; $\widehat{se}$: sensitivity in the samples; $\widehat{sp}$: specificity in the samples; N_0: size of definition-negative persons in the population; N_1: size of definition-positive persons in the population; PPV: positive predictive value in the population; NPV: negative predictive value in the population; SE: sensitivity in the population; SP: specificity in the population; n_1: size of

positive samples; se_0: value of se^* that maximizes n_0 in Equation (7).

Table 2 Validation studies with stratified sampling where the size of negative samples (n_0) is specified: Monte Carlo simulations with 10,000 iterations

Study ID	Option	$n_0{}^*$	δ_{se_max}	$\delta_{se}(0.7)$	$\widehat{PPV}$	$\widehat{NPV}$	$\widehat{se}$	$\widehat{sp}$
N_0/N_1=24, PPV=0.8, NPV=0.99, SE=0.769, SP=0.992, n_1=100 (Study ID= 28-30)								
28	D	150	0.346 (0.326-0.369) [a]	0.344	0.800 (0.720-0.880) [a]	0.993 (0.973-1.000) [a]	0.828 (0.546-1.000) [a]	0.992 (0.988-0.995) [a]
29	D	300	0.231 (0.218-0.246)	0.229	0.800 (0.720-0.880)	0.990 (0.977-1.000)	0.774 (0.591-1.000)	0.992 (0.988-0.995)
30	D	600	0.157 (0.149-0.168)	0.156	0.800 (0.720-0.880)	0.990 (0.982-0.998)	0.769 (0.649-0.952)	0.992 (0.988-0.995)
N_0/N_1=24, PPV=0.8, NPV=0.98, SE=0.625, SP=0.992, n_1=100 (Study ID= 31-33)								
31	D	150	0.346 (0.326-0.369)	0.344	0.800 (0.720-0.880)	0.980 (0.953-1.000)	0.631 (0.422-1.000)	0.992 (0.988-0.995)
32	D	300	0.231 (0.218-0.246)	0.229	0.800 (0.720-0.880)	0.980 (0.963-0.993)	0.631 (0.470-0.838)	0.992 (0.988-0.995)
33	D	600	0.157 (0.149-0.168)	0.156	0.800 (0.720-0.880)	0.980 (0.968-0.992)	0.625 (0.516-0.794)	0.992 (0.988-0.995)
N_0/N_1=24, PPV=0.8, NPV=0.96, SE=0.425, SP=0.991, n_1=100 (Study ID= 34-36)								
34	D	150	0.346 (0.326-0.369)	0.344	0.800 (0.720-0.880)	0.960 (0.927-0.987)	0.461 (0.307-0.722)	0.991 (0.988-0.995)
35	D	300	0.231 (0.218-0.246)	0.229	0.800 (0.720-0.880)	0.960 (0.937-0.980)	0.458 (0.342-0.631)	0.991 (0.988-0.995)
36	D	600	0.157 (0.149-0.168)	0.156	0.800 (0.720-0.880)	0.960 (0.945-0.975)	0.455 (0.371-0.579)	0.991 (0.988-0.995)

a Medians and 2.5th-97.5th percentile range (in parentheses).

Option: option to select the size of negative samples (see text for the details); $n_0{}^*$: size of negative samples arbitrarily selected by the investigator; δ_{se_max}: maximum possible precision in Equation (9); $\delta_{se}(0.7)$: approximated maximum precision in Equation (10); $\widehat{PPV}$: positive predictive value in the

samples; $\widehat{NPV}$: negative predictive value in the samples; $\widehat{se}$: sensitivity in the samples; $\widehat{sp}$: specificity in the samples; N_0: size of definition-negative persons in the population N_1: size of definition-positive persons in the population; PPV: positive predictive value in the population; NPV: negative predictive value in the population SE: sensitivity in the population; SP: specificity in the population; n_1: size of positive samples.

Table 3 95% confidence interval of the sensitivity ($95\%CI_{se}$) and coverage: Monte Carlo simulations with 10,000 iterations

Study ID	$\delta_{se}{}^{*}$	δ_{se_max}	Equation (1)			Equation (5)			Bootstrap method		
			$\widehat{se}$	δ_{se}	coverage	$\widehat{se}$	δ_{se}	coverage	$\widehat{se}$	δ_{se}	coverage
1	0.2	-	0.764 (0.607-0.963) [a]	0.189 [b]	0.893 [c]	0.764 (0.566-0.963) [d]	0.189 [b]	0.893 [c]	0.773 (0.606-0.937) [e]	0.181 [f]	0.900 [g]
4	0.15	-	0.760 (0.635-0.909)	0.148	0.937	0.760 (0.610-0.909)	0.148	0.937	0.768 (0.632-0.916)	0.139	0.952
7	0.1	-	0.771 (0.683-0.870)	0.099	0.927	0.771 (0.671-0.870)	0.099	0.955	0.773 (0.673-0.869)	0.096	0.951
10	0.2	-	0.618 (0.475-0.806)	0.187	0.931	0.618 (0.431-0.806)	0.187	0.963	0.630 (0.474-0.821)	0.188	0.941
13	0.15	-	0.627 (0.511-0.768)	0.142	0.930	0.627 (0.485-0.768)	0.142	0.957	0.631 (0.502-0.777)	0.144	0.947
16	0.1	-	0.625 (0.543-0.719)	0.095	0.944	0.625 (0.530-0.719)	0.095	0.961	0.626 (0.529-0.722)	0.100	0.962
19	0.2	-	0.449 (0.341-0.590)	0.141	0.948	0.449 (0.308-0.590)	0.141	0.963	0.463 (0.342-0.619)	0.158	0.954
22	0.15	-	0.456 (0.369-0.563)	0.107	0.941	0.456 (0.349-0.563)	0.107	0.966	0.458 (0.358-0.576)	0.120	0.964
25	0.1	-	0.454 (0.391-0.527)	0.073	0.950	0.454 (0.381-0.527)	0.073	0.965	0.454 (0.376-0.536)	0.081	0.973
28	-	0.346	0.828 (0.591-1.000)	0.172	0.991	0.828 (0.497-1.000)	0.172	0.991	0.833 (0.595-1.000)	0.167	0.778
29	-	0.231	0.774 (0.599-0.999)	0.218	0.933	0.774 (0.548-0.999)	0.218	0.964	0.791 (0.596-1.000)	0.194	0.936
30	-	0.156	0.769 (0.639-0.926)	0.154	0.919	0.769 (0.613-0.926)	0.154	0.922	0.776 (0.633-0.922)	0.150	0.924
31	-	0.346	0.631 (0.417-0.955)	0.286	0.956	0.631 (0.306-0.955)	0.286	0.961	0.644 (0.434-1.000)	0.279	0.929
32	-	0.231	0.631 (0.470-0.847)	0.216	0.927	0.631 (0.411-0.847)	0.216	0.957	0.639 (0.455-0.906)	0.230	0.943
33	-	0.156	0.625 (0.505-0.774)	0.148	0.934	0.625 (0.476-0.774)	0.148	0.957	0.629 (0.497-0.778)	0.153	0.951
34	-	0.346	0.461 (0.300-0.705)	0.245	0.926	0.461 (0.213-0.705)	0.245	0.966	0.474 (0.310-0.744)	0.274	0.937
35	-	0.231	0.458 (0.337-0.623)	0.164	0.941	0.458 (0.294-0.623)	0.164	0.971	0.461 (0.335-0.679)	0.218	0.967
36	-	0.156	0.455 (0.365-0.567)	0.112	0.946	0.455 (0.342-0.567)	0.112	0.968	0.455 (0.361-0.589)	0.134	0.965

a Median of $\widehat{se}$ from Equation (3) and medians of the upper and lower limits of $95\%CI_{se}$ from Equation (1) (in parentheses).

b Median of δ_{se} in Equation (5) or 1-$\widehat{se}$ when the upper limit of $\widehat{se}$ is set to 1.

c Proportion of 10,000 iterations in which $95\%CI_{se}$ includes SE.

d Median of $\widehat{se}$ from Equation (3) and medians of the upper and lower limits of $95\%CI_{se}$ from Equation (5) (in parentheses).

e Median of the 50th percentiles of $\widehat{se}$ by the bootstrap method and medians of the 2.5th and 97.5th percentiles of $\widehat{se}$ by the bootstrap method (in parentheses).

f Median of the difference between the 97.5th and 50th percentiles of $\widehat{se}$ among 1,000 resamples.

g Proportion of 10,000 iterations in which the 2.5th-97.5th percentile range of 1,000 resamples includes SE.

$\delta_{se}{}^{*}$: intended precision of the sensitivity; δ_{se_max}: maximum possible precision from Equation (9); $\widehat{se}$: sensitivity in the samples; δ_{se}: the value of δ_{se} from Equation (5) or 1-$\widehat{se}$ when the upper limit of $\widehat{se}$ is set to 1 and the difference between the 97.5th and 50th percentiles of $\widehat{se}$ by the bootstrap method.

Table 4 95% confidence interval of the specificity ($95\%CI_{sp}$) and coverage: Monte Carlo simulations with 10,000 iterations

Study ID	Equation (2)			Equation (6)			bootstrap method		
	$\widehat{sp}$	δ_{sp}	coverage	$\widehat{sp}$	δ_{sp}	coverage	$\widehat{sp}$	δ_{sp}	coverage
1	0.9917 (0.9884-0.9949) [a]	0.0032 [b]	0.933 [c]	0.9917 (0.9884-0.9949) [d]	0.0032 [b]	0.933 [c]	0.9915 (0.9877-0.9951) [e]	0.0034 [f]	0.968 [g]
4	0.9917 (0.9884-0.9949)	0.0032	0.932	0.9917 (0.9884-0.9949)	0.0032	0.932	0.9916 (0.9876-0.9951)	0.0033	0.961
7	0.9917 (0.9884-0.9949)	0.0032	0.932	0.9917 (0.9884-0.9949)	0.0032	0.932	0.9917 (0.9881-0.9954)	0.0033	0.968
10	0.9916 (0.9883-0.9949)	0.0033	0.931	0.9916 (0.9883-0.9949)	0.0033	0.931	0.9915 (0.9876-0.9950)	0.0035	0.966
13	0.9916 (0.9883-0.9949)	0.0033	0.932	0.9916 (0.9883-0.9949)	0.0033	0.932	0.9916 (0.9875-0.9950)	0.0034	0.964
16	0.9916 (0.9883-0.9949)	0.0033	0.927	0.9916 (0.9883-0.9949)	0.0033	0.927	0.9916 (0.9879-0.9953)	0.0034	0.964
19	0.9915 (0.9881-0.9948)	0.0033	0.931	0.9915 (0.9881-0.9948)	0.0033	0.931	0.9913 (0.9874-0.9949)	0.0035	0.964
22	0.9914 (0.9881-0.9948)	0.0033	0.928	0.9914 (0.9881-0.9948)	0.0033	0.929	0.9914 (0.9872-0.9949)	0.0035	0.958
25	0.9914 (0.9881-0.9948)	0.0033	0.927	0.9914 (0.9881-0.9948)	0.0033	0.927	0.9915 (0.9877-0.9952)	0.0034	0.963
28	0.9917 (0.9884-0.9949)	0.0032	0.932	0.9917 (0.9884-0.9949)	0.0032	0.932	0.9917 (0.9873-0.9952)	0.0035	0.969
29	0.9917 (0.9885-0.9949)	0.0032	0.931	0.9917 (0.9884-0.9949)	0.0032	0.931	0.9918 (0.9881-0.9952)	0.0033	0.966
30	0.9917 (0.9885-0.9949)	0.0032	0.932	0.9917 (0.9884-0.9949)	0.0032	0.932	0.9917 (0.9876-0.9951)	0.0034	0.970
31	0.9916 (0.9884-0.9949)	0.0033	0.934	0.9916 (0.9884-0.9949)	0.0033	0.934	0.9916 (0.9872-0.9951)	0.0036	0.970
32	0.9916 (0.9883-0.9949)	0.0033	0.932	0.9916 (0.9883-0.9949)	0.0033	0.932	0.9917 (0.9880-0.9951)	0.0034	0.965
33	0.9916 (0.9883-0.9949)	0.0033	0.932	0.9916 (0.9883-0.9949)	0.0033	0.932	0.9917 (0.9875-0.9951)	0.0034	0.970
34	0.9915 (0.9881-0.9948)	0.0033	0.935	0.9915 (0.9881-0.9948)	0.0033	0.935	0.9915 (0.9870-0.9951)	0.0037	0.970
35	0.9915 (0.9881-0.9948)	0.0033	0.934	0.9915 (0.9881-0.9948)	0.0033	0.934	0.9916 (0.9877-0.9950)	0.0035	0.965
36	0.9914 (0.9881-0.9948)	0.0033	0.929	0.9914 (0.9881-0.9948)	0.0033	0.929	0.9915 (0.9872-0.9950)	0.0035	0.966

a Median of $\widehat{sp}$ from Equation (4) and medians of the upper and lower limits of $95\%CI_{sp}$ from Equation (2) (in parentheses).

b Median of δ_{sp} from Equation (6).

c Proportion of 10,000 iterations in which $95\%CI_{sp}$ includes SP.

d Median of $\widehat{sp}$ from Equation (4) and medians of the upper and lower limits of $95\%CI_{sp}$ from Equation (6) (in parentheses).

e Median of the 50th percentile of $\widehat{sp}$ by the bootstrap method and medians of the 2.5th and 97.5th percentiles of $\widehat{sp}$ by the bootstrap method (in parentheses).

f Median of the difference between the 97.5th and 50th percentiles of $\widehat{sp}$ in 1,000 resamples.

g Proportion of 10,000 iterations in which the 2.5th-97.5th percentile range of 1,000 resamples includes SP.

$\widehat{sp}$: specificity in the samples; δ_{sp}: the value of δ_{sp} from Equation (6) and difference between the 97.5th and 50th percentiles of $\widehat{sp}$ by the bootstrap method.

Table 5 Coverage of $95\%CI_{se}$ and $95\%CI_{sp}$: Post hoc studies

Coverage	$95\%CI_{se}$		$95\%CI_{sp}$	
	Equation (1)	Equation (5)	Equation (2)	Equation (6)
	N (%)	N (%)	N (%)	N (%)
≥0.95	4,069 (38.5%)	8,887 (84.2%)	267 (2.5%)	355 (3.4%)
0.93-0.95	4,688 (44.4%)	895 (8.5%)	8,335 (78.9%)	8,292 (78.5%)
<0.93	1,803 (17.1%)	778 (7.4%)	1.958 (18.5%)	1,913 (18.1%)
Total	10,560 (100%)	10,560 (100%)	10,560 (100%)	10,560 (100%)

The coverage of $95\%CI_{se}$ (proportion of studies in which $95\%CI_{se}$ includes SE) and coverage of $95\%CI_{sp}$ (proportion of studies in which $95\%CI_{sp}$ includes SP) are shown. Each of the 10,560 studies was conducted with one of 10,560 combinations of N_0/N_1 (9, 19, 24, 49, 99), PPV (0.5, 0.6, 0.7, 0.8, 0.9, 0.95), NPV (0.8, 0.9, 0.93, 0.96, 0.98, 0.99, 0.995, 0.9975), $\delta_{se}{}^{*}$ (0.1, 0.15, 0.2, 0.25) and n_1 (60, 80, 100, 120, 150, 200, 300, 500, 800, 1000, 20000).

Additional file

Additional file 1: Derivation of equations in the text (DOCX 934 kb)

List of abbreviations

CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value; RA: rheumatoid arthritis

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Funding

There is no funding source to declare.

Authors' contributions

KK conceptualized this study and was responsible for the methodology and analysis and for writing the initial draft. MI and TY were involved in the review and editing. All authors have read and reviewed the final manuscript.

Acknowledgements

The authors wish to thank the “Task force on validation studies of outcome definition in claims data in Japan” (the report (in Japanese) is available online at http://www.jspe.jp/committee/pdf/validationtrr120180528.pdf), of which two authors of the current manuscript (KK and MI) are members. The task force was established by the Japanese Society for Pharmacoepidemiology, and the activity of the task force motivated the current study.

Competing interests

The authors declare that they have no competing interests.

References

1. Ando T, Ooba N, Mochizuki M, Koide D, Kimura K, Lee SL, et al. Positive predictive value of ICD-10 codes for acute myocardial infarction in Japan: a validation study at a single center. BMC Health Services Research 2018;18:895.
2. McGuinness LA, Warren-Gash C, Moorhouse LR, Thomas SL. The validity of dementia diagnoses in routinely collected electronic health records in the United Kingdom: A systematic review. Pharmacoepidemiol Drug Saf 2019; 28:244-55.
3. Cutrona SL, Toh S, Iyer A, Foy S, Cavagnaro E, Forrow S, et al. Design for Validation of Acute Myocardial Infarction Cases in Mini-Sentinel. Pharmacoepidemiol Drug Saf 2012; 21: 274-81.
4. Widdifield J, Bombardier C, Bernatsky S, Paterson JM, Green D, Young J, et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. BMC Musculoskelet Disord. 2014;15:216. doi: 10.1186/1471-2474-15-216.
5. Husain N, Blais P, Kramer J, Kowalkowski M, Richardson P, El-Serag HB, et al. Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data. Aliment Pharmacol Ther. 2014;40:949-954. doi: 10.1111/apt.12923.
6. Klungsøyr K, Harmon QE, Skard LB, Simonsen I, Austvoll ET, Alsaker ER, et al. Validity of pre-eclampsia registration in the medical birth registry of norway for women participating in the norwegian mother and child cohort study, 1999-2010. Paediatr Perinat Epidemiol 2014;28(5):362-371. doi: 10.1111/ppe.12138.
7. Katz D, Baptista J, Azen SP, Pike MC. Obtaining confidence intervals for the risk ratio in cohort studies. Biometrics 1978; 34: 469-474.
8. Walsh KE, Cutrona SL, Foy S, Baker MA, Forrow S, Shoaibi A, et al. Validation of anaphylaxis in the Food and Drug Administration's Mini-Sentinel. Pharmacoepidemiol Drug Saf 2013;22(11):1205-13. doi: 10.1002/pds.3505.
9. Lacasse Y, Daigle JM, Martin S, Maltais F. Validity of chronic obstructive pulmonary disease diagnoses in a large administrative database. Can Respir J 2012;19(2):e5-9.
10. Goldberg DS, Lewis JD, Halpern SD, Weiner MG, Lo Re V, 3rd. Validation of a coding algorithm to identify patients with hepatocellular carcinoma in an administrative database. Pharmacoepidemiol Dug Saf 2013;22(1):103-7. doi: 10.1002/pds.3367.
11. Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything alright? JAMA 1983;259:1743–5.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FYMGLFV41462C8C1EDB541192B85.pdf
- AdditionalFile1.pdf