Errors Not Yet Found Can Be Counted from Two Readers' Overlap, but Come Out Too Few If Both Miss the Same Ones

— If two readers find **20** and **25** errors with **10** in common, the total is estimated at **50.000000** and the errors not yet found at **15.000000** — if the readers' blind spots are opposite, **27** are reported as **155.571429** — [Paper 585]

Yuuki YAMAGISHI (山岸 勇輝)
Yamagishi Complexity Torus Lab, Japan
September 2026

— From thinking to syncing. —

Abstract

One way to estimate how many errors remain in a manuscript is to have two people read it separately and use the overlap of what they found.What this paper shows is that the estimate is right only when the two readers' misses are unrelated; if they miss the same errors it comes out too few, if they miss opposite ones too many, and which way it errs is set by the sign of the covariance of their misses.No new theorem or law is claimed. First, the overlap gives the total——if A finds 20, B finds 25 and both find 10, the total is estimated at 50.000000 and the errors not yet found at 15.000000; the small-sample correction gives 48.636364 (Section 2). **Second, if findability is uniform, the estimate is exactly right——with 100 true errors the expected estimate is 100.000000, and the estimate of those not found, 30.000000, matches the true 30** (Section 3). **Third, and this is the core.If both readers miss the same errors, the estimate comes out too few——when half the errors are found by each reader with probability 0.7 and half with 0.1, the overlap swells to 25.0000, the total is estimated at 64.000000, and the 45 errors not yet found are reported as 9.000000** (Section 4). **Fourth, if their blind spots are opposite, the estimate comes out too many——the overlap shrinks to 7.0000, the total is estimated at 228.571429, and 27 errors not yet found are reported as 155.571429** (Section 5). **Fifth, the error of the estimate has one formula——the ratio of estimate to truth is $1/(1+\mathrm{cov}/(\bar{p}_1\bar{p}_2))$, matching the direct calculation to 10^{-12}** (Section 6). **Sixth, the separator is whether the two readers' misses are related, and in which direction——a positive covariance gives too few, a negative one too many, and only zero is right** (Section 6). **Keywords**: capture-recapture, Lincoln-Petersen estimator, proofreading, number of errors, independence assumption.

Scope of this paper (scope note): No new theorem or law is claimed——the Lincoln--Petersen estimate and the Chapman correction are standard in capture--recapture.Only expected values are computed——the variation from one trial to another and confidence intervals are not treated.Findability is split into two representative groups——half at 0.7 and half at 0.1; the distribution in real manuscripts is not stated.Only two readers are considered.No proofreading procedure is designed——who should read how many times is not stated.Relation to earlier papers: Paper 334 showed that combining independent measurements adds precisions, but that perfectly correlated sensors are one——here too the assumption of independence decides the number: there correlation broke the addition of precisions, here it breaks the estimated total, and by its direction the estimate errs either way.Paper 579 showed that a correct-response rate is not a property of the question alone but a relation between question and group——here too, that findability differs from error to error is what moves the number.Paper 300 showed that sharing a root is decidable——coming out too few and coming out too many share one root, the covariance of the misses.What is added is confirming by expected values that the estimate is exactly right when findability is uniform, setting out that shared misses report 45 unfound errors as $\mathbf{9.000000}$ and opposite misses report 27 as $\mathbf{155.571429}$, placing the error in a covariance formula checked against the direct calculation, and placing the separator on the sign of the covariance.

1. Notation and the Question

Setting: a manuscript with N errors is read separately by A and B; A finds n_1, B finds n_2 and both find m.If their misses are unrelated, the fraction A found among all errors equals the fraction A found among those B found, so

$$\hat{N} = \frac{n_1 \, n_2}{m}$$

estimates the total (Lincoln--Petersen).The estimate of errors not yet found is $\hat{N} - (n_1 + n_2 - m)$.

Observation 1.1

errors not yet found seem uncountable, being unseen.Yet a large overlap between two readers suggests few errors, and a small one many——the same reasoning estimates the fish in a pond from how many marked and released fish reappear in a second catch.

Question: when is that estimate right, and when does it err, and which way?

2. Two Readers

Computation 2.1: A finds 20 errors, B finds 25, both find 10.The total is estimated at $\mathbf{50.000000}$; 35 were found, so $\mathbf{15.000000}$ are estimated not yet found.The small-sample correction (Chapman) gives $\mathbf{48.636364}$.

> **Observation 2.2**
>
> **the number of the unseen comes from the overlap of the seen.If B also found half of what A found ($10/20$), B is taken to have found half of everything, and the whole is twice 25, or 50——this reasoning assumes that the two miss errors separately, without relation to each other.**

3. If Findability Is Uniform

Computation 3.1: **a true total of 100, with every error found by A with probability 0.4 and by B with 0.5.In expectation $n_1 = \mathbf{40.0000}$, $n_2 = \mathbf{50.0000}$, overlap $\mathbf{20.0000}$, and the estimated total is $\mathbf{100.000000}$ (matching the true 100 to 10^{-9}).**

> **Observation 3.2**
>
> **if findability is uniform, the estimate is exactly right.$\mathbf{70.0000}$ are found, and the estimate of those not found, $\mathbf{30.000000}$, matches the true 30——the two readers may differ in skill; as long as findability does not differ from error to error, the estimate does not err.**

4. When Both Miss the Same Errors

Computation 4.1: **of 100 true errors, half are found by both readers with probability 0.7 and half by both with only 0.1.**

quantity	expected value
found by A	**40.0000**
found by B	**40.0000**
overlap	**25.0000**
estimated total	**64.000000**
found	**55.0000**
not yet found (estimated)	**9.000000**
not yet found (true)	**45.000000**

> **Observation 4.2**
>
> **this is the core.When both readers miss the same errors, the 45 not yet found are reported as 9——the easy errors are found by both and swell the overlap, while the hard ones are missed by both and appear neither in the overlap nor in the counts. A large overlap is misread as “almost all found”.**

5. When They Miss Opposite Errors

Computation 5.1: **the same two groups, but A finds the first half with 0.7 and the second with 0.1, while B does the reverse.**

quantity	expected value
overlap	**7.0000**
estimated total	**228.571429**
found	**73.0000**
not yet found (estimated)	**155.571429**
not yet found (true)	**27.000000**

Observation 5.2

with opposite blind spots, the estimate comes out too many.Each reader covers the other's misses, the overlap shrinks, and 27 unfound errors are reported as 155.571429——the same formula errs in either direction.

6. The Formula of the Error, and the Separator

Computation 6.1: **from the mean findabilities $\bar{p}_1, \bar{p}_2$ and the covariance cov of findability across errors, the ratio of estimate to truth is**

$$\frac{\hat{N}}{N} = \frac{1}{1 + \mathrm{cov}/(\bar{p}_1\,\bar{p}_2)}.$$

For the group that misses the same errors (covariance 0.0900) and for the group that misses opposite ones (covariance -0.0900), it matches the direct calculation to 10^{-12}.

Observation 6.2

the separator is the sign of the covariance of the two readers' misses.A positive covariance (missing the same errors) gives too few, a negative one (missing opposite errors) too many, and only zero is right——whether the estimate is right cannot be told from the counts found; it is decided by an unseen property, whether the two are weak on the same kinds of error.

Observation 6.3

as in Paper 334, the assumption of independence decides the number.There, combining independent measurements added precisions, but perfectly correlated sensors were one——here correlation breaks the estimated total, and by its direction the estimate errs either way.

Observation 6.4

as in Paper 579, non-uniformity moves the number.There, questions of different discrimination reversed their order of correct-response rate between groups——here, errors of different findability, shared between the two readers, make the estimate come out too few.

Observation 6.5

one root, under Paper 300.Coming out too few and coming out too many come from one root, the covariance of the misses——only the sign differs.

Observation 6.6

what is not said.No proofreading procedure is designed——what is shown is only on what assumption an estimate from the overlap is right, and which way it errs when the assumption fails.

7. Roll Call and Closing

1. **20, 25, overlap 10: estimate 50.000000** (Computation 2.1): **15.000000 not yet found.**
2. **small-sample correction 48.636364** (Computation 2.1): **Chapman.**
3. **uniform findability: estimate 100.000000** (Computation 3.1): **exactly right.**
4. **shared misses swell the overlap to 25.0000** (Computation 4.1): **it swells.**
5. **45 unfound reported as 9.000000** (Observation 4.2): **the core.**
6. **opposite misses shrink the overlap to 7.0000** (Computation 5.1): **it shrinks.**
7. **27 reported as 155.571429** (Observation 5.2): **too many.**
8. **the error is $1/(1+\mathrm{cov}/(\bar{p}_1\bar{p}_2))$** (Computation 6.1): **matching the direct calculation.**
9. **the separator is the sign of the covariance** (Observation 6.2): **positive too few, negative too many.**
10. **independence decides, as in Paper 334** (Observation 6.3): **correlation sets the number.**

Closing: **One way to estimate how many errors remain in a manuscript is to have two people read it separately and use the overlap of what they found.The number of the unseen comes from the overlap of the seen——if A finds 20, B 25 and both 10, the total is estimated at 50.000000 and 15.000000 remain, or 48.636364 with the small-sample correction; with uniform findability the expected estimate hits the true 100 exactly.But if both readers miss the same errors, the estimate comes out too few——with half the errors easy for both and half hard for both, the overlap swells to 25.0000, the total is estimated at 64.000000, and the 45 not yet found are reported as 9.000000: a large overlap is misread as "almost all found". With opposite blind spots the overlap shrinks to 7.0000 and 27 unfound errors are reported as 155.571429.The separator is the sign of the covariance of the two readers' misses——the ratio of estimate to truth is $1/(1+\mathrm{cov}/(\bar{p}_1\bar{p}_2))$: too few for a positive covariance, too many for a negative one, right only at zero. Whether the estimate is right cannot be told from the counts; it rests on an unseen property, whether**

the two are weak on the same errors. Under Paper 300 the two directions of error share one root.Placed among the earlier papers——as in Paper 334 independence decides the number, and here its direction makes the estimate err either way. As in Paper 579, non-uniformity moves the number.To be honest——only expected values were computed, with no variation or confidence intervals; findability was split into two representative groups and only two readers considered. And no proofreading procedure is designed: what is shown is on what assumption the estimate is right, and which way it errs.

References

1. Chapman, D. G. (1951). *Some Properties of the Hypergeometric Distribution with Applications to Zoological Sample Censuses*. University of California Publications in Statistics.
2. Lincoln, F. C. (1930). *Calculating Waterfowl Abundance on the Basis of Banding Returns*. United States Department of Agriculture Circular.
3. Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters* (2nd ed.). Griffin.
4. Yamagishi, Y. (2026). *The Three Hundredth Census: Whether Two Things Share a Root Is Decidable* [Paper 300]. Zenodo. 10.5281/zenodo.22171914
5. Yamagishi, Y. (2026). *What Is Being Added Is Not the Values but the Precisions* [Paper 334]. Zenodo. 10.5281/zenodo.22681804
6. Yamagishi, Y. (2026). *A Correct-Response Rate Is Not a Property of the Question but a Relation Between Question and Group* [Paper 579]. Zenodo. 10.5281/zenodo.23014333

— From thinking to syncing. —

Yuuki YAMAGISHI (山岸 勇輝)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)