

A Gap That Creates No Bias Still Moves the Sign

— Two gaps filled on the same day by the same procedure cut in opposite directions — one took a mean of $+0.45$ to -2.22 , the other took $+4.2$ to $+106.8$ — the 15 points that were filled in average -10.41 , almost the same magnitude as the overall mean with the opposite sign — [Paper 420]

Yuuki YAMAGISHI (山岸 勇輝)

Yamagishi Complexity Torus Lab, Japan

September 2026

— *From thinking to syncing.* —

Abstract

Missing data is sometimes treated as harmless when it introduces no bias. This paper shows that the sign moves even when there is no bias—and that until the gap is filled, which way it will move cannot be told. No new theorem or law is claimed. First, this is the core of the paper: Two gaps filled on the same day by the same procedure cut in opposite directions: one moved a mean from $+0.45$ to -2.22 , the other from $+4.2$ to $+106.8$ (Section 3). Second, the mean of the 15 filled points alone is fixed by inversion at -10.41 , almost the same magnitude as the overall mean of $+10.19$ with the opposite sign (Section 2). Third, the inversion is confirmed independently: the change in the overall total, -154.05 , and the change in the interval, -156.12 , agree to within rounding (Section 2). Fourth, the interval changed sign while the whole did not ($+10.19 \rightarrow +9.88$). Magnitude and sign are separate questions (Section 2). Fifth, dropping points at random moves the mean as well: in a synthetic case $+0.3299$ became $+0.6643$ (Section 4). Sixth, the separator is not whether there is bias but whether the gap can be filled (Section 4).

Scope of this paper (scope note): No new theorem or law is claimed—the classification of missingness and the general theory of imputation are all standard. No imputation method is proposed—whether multiple imputation, maximum likelihood or inverse-probability weighting is best is not the subject. Only “what happened when the same gap was filled from another source” is treated. No particular field is discussed—the figures come from observations of a time series, but neither the market nor any method is the subject, and the quantity is treated as a dimensionless per-item value. Causes are not discussed—why the gap arose is not the subject. No property of an estimator is proved—unbiasedness and consistency are not argued. Neither value is declared correct—this paper has no ground for saying the filled version is right. What is shown is only that filling moves the answer. Relation to earlier papers: Paper 419 showed that the same material and shape can differ in strength by orientation—there a quantity absent from the drawing did the work; here points absent from the record do. Paper 300 showed that common root and distinct root can be decided—“biased missingness” and “unbiased missingness” are distinct roots by that criterion. Paper 412 showed that “nearest” is not defined until the distance is—here a “mean” is not defined until the sample is. Paper 364 showed that selection alone raises the mean—there selecting moved it; here being absent does. What is added is fixing the mean of the filled points uniquely by inversion, confirming that inversion independently against the overall total, setting two gaps of the same day that cut opposite ways side by side, separating the sign of the interval from the sign of the whole, and placing the separator at “fillable” rather than “biased”.

1. Notation and the Question

Setting: Suppose n_0 points were observed over an interval with mean m_0 . Later k points are supplied from another source, giving $n_1 = n_0 + k$ points with mean m_1 .

$$m_1 = \frac{n_0 m_0 + k \bar{x}_{\text{add}}}{n_0 + k} \implies \bar{x}_{\text{add}} = \frac{n_1 m_1 - n_0 m_0}{k}$$

The mean of the supplied points alone is fixed by the observations.

Observation 1.1

Missing data is sometimes called harmless when it carries no bias. If points fell out at random the mean should not move—and as a statement about expectations that is correct.

Observation 1.2

But what one holds is not an expectation; it is one sample. In one sample the mean moves even under random dropping—and which way is decided by what the dropped points were.

Question: By how much, and can the direction be anticipated.

2. The Mean of the Filled Points Is Fixed by Inversion

Computation 2.1: The record for one interval.

Sample	Points	Mean
With the gap	46	+0.4500
After filling	61	−2.2200

The supplied points number $61 - 46 = 15$, and inversion gives

$$\bar{x}_{\text{add}} = \frac{61 \times (-2.22) - 46 \times 0.45}{15} = -10.4080$$

Observation 2.2

No assumption enters.It follows algebraically from two counts and two means—and recombining returns −2.220000000.

Observation 2.3

−10.41 is not small.The overall mean of the same sample is +10.19—almost the same magnitude, with the opposite sign.

Computation 2.4: Confirm independently. The whole went from 975 points at +10.19 to 990 points at +9.88.

Quantity	Value
Change in the overall total	−154.05
Change in the interval total	−156.12
Difference	2.07

Observation 2.5

The two routes agree to within rounding. The increments in count agree too: $990 - 975 = 15$ and $61 - 46 = 15$ —an independent confirmation that the inversion is right.

Observation 2.6

The interval changed sign; the whole did not. $+10.19 \rightarrow +9.88$ —magnitude and sign are separate questions, and mixing them lets one call the same gap either fatal or negligible.

3. Two Gaps of the Same Day Cut Opposite Ways

Computation 3.1: On the same day, from the same source, by the same procedure, a gap in another series was also filled.

Series	Before	After	Direction
First (interval mean)	+0.45	−2.22	worse; sign reversed
Second (interval total)	+4.2	+106.8	much better

What was supplied in the second case is $106.8 - 4.2 = +102.6$.

Observation 3.2

The procedure was the same. Same day, same source, same rule—and still one cut downward and the other upward.

Observation 3.3

So “filling makes things worse” is not the lesson either. It moves towards optimism as readily as towards pessimism—there is nothing to predict the direction from before filling.

Observation 3.4

Where the two sources overlap they agree. The median difference over the 46 common points is 0.00—so it is not a difference of source or of computation; the absent points themselves did the moving.

Observation 3.5

That 0.00 is doing work. Had the overlap disagreed, the question would become which source is right—since it does not, the question narrows to “what were the absent points”.

4. The Separator

The separator: not whether there is bias, but whether the gap can be filled.

Computation 4.1: Check synthetically. Drop 15 of 60 points at random.

The mean moves from +0.3299 to +0.6643, a change of 0.3343.

Observation 4.2

There is no bias in the dropping. Points are dropped with uniform probability—and the mean still moves. That the expectation does not move is a different statement from that this sample does not.

Observation 4.3

How far it moves is set by the number dropped and the dispersion. It is of the order of a standard error—so the more dispersed the quantity, the more a few absences matter.

Observation 4.4

And the sign reverses when the mean is near zero. The interval mean +0.45 was near zero—the overall mean +10.19 was far from it and did not reverse.

Observation 4.5

So “no bias, no problem” conflates two things. That an estimator is unbiased and that this sample’s value will not move are different—distinct roots by the criterion of Paper 300, one a property of the procedure, one of a single realisation.

Observation 4.6

To be honest—this paper has no ground for saying the filled version is right. The supplying source can be wrong too—what is shown is that filling moves the answer, not that the filled value is true.

Observation 4.7

One last thing—reporting that a gap exists is not enough. “Thirty per cent is missing” tells the reader neither the direction nor the size—fill what can be filled, show the answer before and after, and always print the counts.

5. Roll Call and Closing

1. 15 points were supplied (Computation 2.1): **that is 61 – 46.**
2. Their mean is –10.4080 (Computation 2.1): **fixed by inversion.**
3. Recombining returns –2.220000000 (Observation 2.2): **a self-test of the inversion.**
4. Overall –154.05 against interval –156.12 (Computation 2.4): **independent agreement.**
5. Counts agree too, $15 = 15$ (Observation 2.5): **two routes.**
6. The interval reversed; the whole went $+10.19 \rightarrow +9.88$ (Observation 2.6): **sign and magnitude differ.**
7. The other gap went $+4.2 \rightarrow +106.8$ (Computation 3.1): **cut the opposite way.**
8. The overlap differs by a median of 0.00 (Observation 3.4): **not a difference of source.**
9. Random dropping gives $+0.3299 \rightarrow +0.6643$ (Computation 4.1): **no bias, still moves.**
10. The separator is bias or fillability (Observation 4.5): **distinct roots by Paper 300.**

Closing: Missing data is sometimes treated as harmless when it carries no bias—if points fell out at random the mean should not move, and as a statement about expectations that is correct. But what one holds is not an expectation; it is one sample, and in one sample the mean moves even under random dropping. How far it moves is

fixed by inversion—when an interval went from 46 points at +0.45 to 61 points at −2.22, the mean of the 15 supplied points alone is −10.4080. No assumption enters; it follows algebraically from two counts and two means, and recombining returns −2.220000000. That value is not small—the overall mean of the same sample is +10.19, almost the same magnitude with the opposite sign. The inversion is confirmed independently: the change in the overall total, −154.05, matches the change in the interval, −156.12, to within rounding, and the increments in count are both 15. And the direction cannot be anticipated—a second gap, filled on the same day from the same source by the same procedure, cut the other way: +4.2 became +106.8, the supplied part being +102.6. Where the two sources overlap they agree, the median difference over the 46 common points being 0.00—so it is not a difference of source or of computation, and the absent points themselves did the moving. One thing separates them—not whether there is bias, but whether the gap can be filled. Dropping 15 of 60 points at random moves a synthetic mean from +0.3299 to +0.6643, with no bias in the dropping at all. That an estimator is unbiased and that this sample’s value will not move are distinct roots by the criterion of Paper 300, one a property of the procedure and one of a single realisation. The sign reverses only when the mean is near zero: the interval mean +0.45 was, and the overall mean +10.19 was not. To be honest—this paper has no ground for saying the filled version is right; the supplying source can be wrong too. What is shown is that filling moves the answer, not that the filled value is true. One last thing—reporting that a gap exists is not enough. “Thirty per cent is missing” tells the reader neither direction nor size. Fill what can be filled, show the answer before and after, and always print the counts.

References

1. Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*.
 2. Little, R. J. A., & Rubin, D. B. (2019). *Statistical Analysis with Missing Data* (3rd ed.). Hoboken: Wiley.
 3. Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
 4. Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods*.
 5. Yamagishi, Y. (2026). *The Three Hundredth Census: Whether Two Things Share a Root Is Decidable* [Paper 300]. Zenodo. 10.5281/zenodo.22171914
 6. Yamagishi, Y. (2026). *Selection Alone Raises the Mean by 1.52 Standard Deviations and Shrinks the Spread to 0.45* [Paper 364]. Zenodo. 10.5281/zenodo.22697530
 7. Yamagishi, Y. (2026). “Nearest” Is Not Defined Until the Distance Is [Paper 412]. Zenodo. 10.5281/zenodo.22699661
 8. Yamagishi, Y. (2026). *The Same Material, the Same Shape, and the Orientation Decides* [Paper 419]. Zenodo. 10.5281/zenodo.22699940
-

— From thinking to syncing. —

Yuuki YAMAGISHI (山岸 勇輝)

