AGRI MAGAZINE
(International E-Magazine for Agricultural Articles)
Volume: 03, Issue: 07 (July, 2026)
Available online at http://www.agrimagazine.in
©Agri Magazine, ISSN: 3048-8656

Missing Data, Missing Farmers: The Hidden Bias in Agricultural Surveys

*Chiranjit Mazumder[1], Mrinmoy Ray[1] and Anbukkani Perumal[2]
[1]ICAR-Indian Agricultural Research Institute, New Delhi, India
[2]ICAR-Indian Institute of Horticultural Research, Bengaluru, India
*Corresponding Author's email: majumder.chira@gmail.com

Key message: A survey may follow an excellent sampling design and still misrepresent agriculture when some farmers are absent from the frame, difficult to contact, unable to answer particular questions, or removed during data cleaning. Missingness becomes bias when the farmers who disappear are systematically different from those who remain.

A farmer can disappear without leaving the field

Imagine two cultivators in the same village. The first owns land recorded in the revenue system, lives near the main road and keeps input receipts. The second farms leased plots through an informal arrangement, moves between the village and a nearby town for wage work, and relies on memory rather than written accounts. Both produce food. Yet the first is far more likely to appear in a list, be found by an investigator, answer every question and survive the final data-cleaning process. The second may vanish statistically while remaining fully active economically.

This is the central danger of missing data in agricultural surveys. The problem is often treated as a technical inconvenience—a few blank cells to be deleted or filled. In reality, missingness can alter who agriculture appears to be composed of, how productive farms seem, who is considered vulnerable and where public programmes are directed. It can turn under-representation into apparently objective evidence.

India's official surveys use scientific designs to move from villages and households to national estimates. For example, the National Sample Survey's 77th Round used a stratified two-stage design and household listing within selected first-stage units (Ministry of Statistics and Programme Implementation [MoSPI], 2021). Such design is indispensable. But design weights cannot fully repair a farmer who was never placed on the frame, a tenant whose cultivation was not disclosed, or a woman whose work was reported by another household member as "help."

Missingness is more than an empty cell

Survey statisticians distinguish several routes through which information is lost. Coverage error occurs before the interview when a target unit is absent, duplicated or incorrectly classified in the sampling frame. Unit nonresponse occurs when an eligible household or holding does not complete the survey. Item nonresponse occurs when the interview is completed but particular questions are unanswered. A fourth problem—especially important in agriculture—is selective listing or measurement error: a respondent reports the main plot but forgets a kitchen garden, reports the household head but omits occasional workers, or rounds local land units inaccurately.

These errors do not have identical consequences. A missing fertilizer expenditure may affect an input-cost estimate. A missing tenant cultivator can affect estimates of farm structure, access to credit, insurance coverage and programme reach simultaneously. The total survey error perspective therefore asks not only, "How large is the sample?" but also, "Who had a real chance of entering the data, and how accurately were their activities represented?" (Groves et al., 2009).

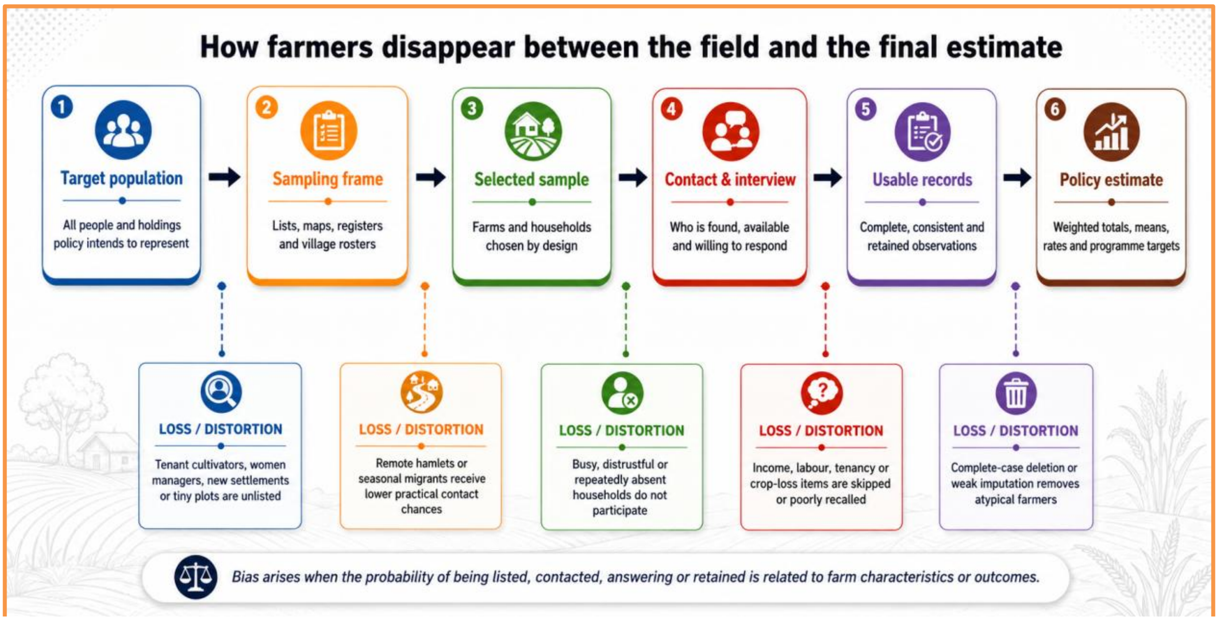

Figure 1. The pathway from the target population to a policy estimate, showing where systematic exclusion can enter.

Table 1. Major forms of missingness in agricultural surveys

Form of error	Typical agricultural example	Who is most exposed?	Likely distortion
Coverage error	A land-register frame omits informal tenants, recent subdivisions or very small plots.	Tenants, sharecroppers, women without title, new or remote settlements	Understates the number and diversity of cultivators; overstates formal ownership.
Unit nonresponse	The selected household is repeatedly absent, refuses, or cannot be reached during fieldwork.	Migrants, wage-dependent households, remote farmers, farmers distrustful of data use	Respondents appear more settled, accessible and institutionally connected than the target population.
Item nonresponse	The household answers crop questions but not debt, income, tenancy or loss questions.	Highly indebted households, informal borrowers, complex mixed-income farms	Biases sensitive indicators and may create false patterns after deletion.
Listing / recall error	Small plots, occasional workers, women's labour or minor crops are not reported.	Marginal plots, secondary workers, women, diversified farms	Changes denominators and can inflate measured productivity or labour efficiency.

Why agricultural surveys are especially vulnerable

Agriculture is difficult to measure because it is seasonal, spatially dispersed and organised through households, plots, livestock units, common resources and informal contracts. A "farm" is not always a single compact unit, and the person who answers the questionnaire may not manage every plot. Good agricultural survey design therefore requires explicit decisions about whether the unit is the household, holding, plot, manager, worker or enterprise (Dillon et al., 2021).

Recall adds another layer. Farmers may be asked after harvest to reconstruct labour, input use, crop loss and sales across several months. Experimental evidence from Ghana shows that end-of-season recall can selectively omit marginal plots and workers; omitted workers were more likely to be women or people for whom farming was not the main occupation. Such omissions changed the composition of the data and affected measured labour productivity (Gaddis et al., 2021). The lesson is not that farmers are unreliable. It is that questionnaires can demand a level of retrospective accounting that many farm systems do not naturally produce.

Missingness is also social. Formal land ownership is easier to verify than informal cultivation. The person introduced as the "farmer" may be male even when a woman manages livestock, seed, post-harvest work or a separate plot. Households in remote hamlets may require more travel time per interview. Migrant and wage-dependent cultivators may be unavailable during standard visiting hours. Each operational shortcut can therefore have a distributional consequence.

When missingness changes the story

Suppose a survey's formally recorded owner-operators respond at much higher rates than tenant cultivators or women managing plots without formal title. Even if every response is internally accurate, the completed sample will no longer resemble the target population. Figure 2 illustrates the mechanism using hypothetical shares and response rates. The better-represented group grows from 45% of the target population to about 53% of completed records, while the least reachable group falls from 10% to below 7%. No observation has been falsified; the bias is created by unequal disappearance.

This matters because agricultural indicators are ratios. Yield is production divided by area. Labour productivity is output divided by labour. Insurance coverage is insured farmers divided by eligible farmers. If small plots, occasional workers or informal cultivators are disproportionately missing, both numerators and denominators can shift. A neat average may therefore conceal a distorted population.

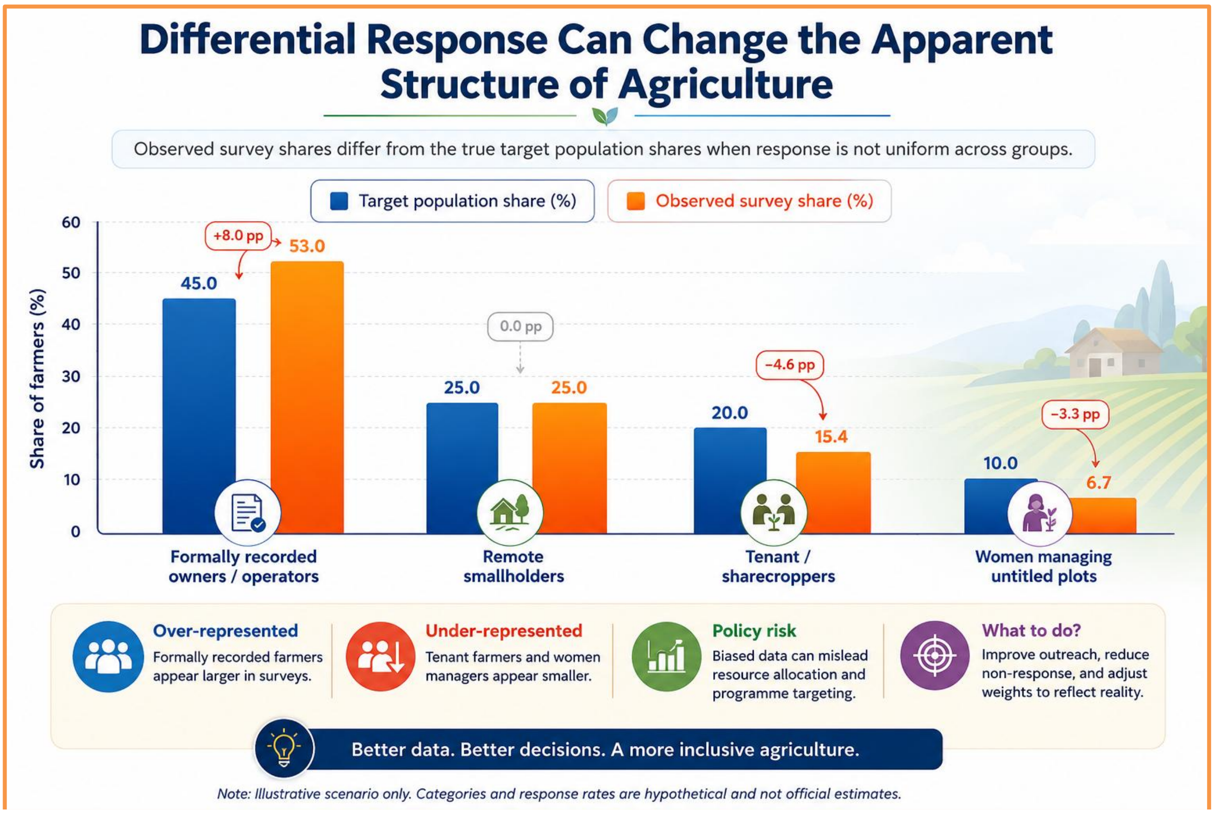

Figure 2. An illustrative example of how unequal response rates alter observed group shares. Values are hypothetical, not official estimates.

Digital agriculture can reduce gaps—or automate them

Digital registries, georeferenced maps, crop-sown records and computer-assisted interviews can improve timeliness, validation and linkage. India's Digital Agriculture Mission combines farmer registries, georeferenced village maps and crop-sown registries within a federated digital infrastructure (Ministry of Agriculture & Farmers Welfare, 2024). This creates a major opportunity to strengthen frames and reduce duplication.

However, digital data are not automatically complete. A registry built primarily around land records may describe legal ownership better than actual cultivation. Smartphone-based contact can miss households with weak connectivity, shared devices or limited digital literacy. Satellite data may identify cropped area but cannot, by itself, identify who controls the plot, who supplied labour or why a farmer did not enrol. Digital systems should therefore be treated as complementary sources that require field verification, transparent inclusion rules and mechanisms for correction—not as unquestionable replacements for surveys.

Table 2. Statistical responses to missing data—and their limits

Method	What it does	Appropriate use	Critical caution
Complete-case analysis	Uses only records with all required variables observed.	Limited missingness plausibly unrelated to outcomes or covariates.	Often discards rare or vulnerable groups and can increase both bias and uncertainty.
Weighting adjustment	Raises the influence of respondents who resemble nonrespondents on known characteristics.	Unit nonresponse when reliable auxiliary variables and design information exist.	Cannot recover people absent from both the frame and auxiliary data; extreme weights reduce precision.
Single imputation	Fills one value using a donor, mean, regression or deterministic rule.	Operational datasets and low-stakes descriptive completion with careful flagging.	Treating filled values as known usually understates uncertainty.
Multiple imputation	Creates several plausible completed datasets and combines estimates.	Item nonresponse when a defensible model includes variables related to missingness and the outcome.	Results depend on the imputation model; it does not solve structural undercoverage.
Sensitivity analysis	Tests conclusions under alternative assumptions about unobserved values or groups.	When missingness may depend on unobserved outcomes or exclusion cannot be ruled out.	Should be reported as a range of credible conclusions, not hidden as a technical appendix.

From "clean data" to credible representation

The first remedy is preventive, not computational. Frames should be built from multiple sources—population lists, land and programme records, village mapping and local verification—and should explicitly search for cultivators who lack formal title. Listing protocols should ask separately about owned, leased-in, leased-out, sharecropped and group-managed land. Plot and worker rosters should use local terminology and prompts for small plots, kitchen gardens, secondary crops and occasional labour.

Second, fieldwork should be designed around farmers' calendars. Revisits, appointments outside standard hours, local-language interviewing and respondent substitution rules that preserve subject knowledge can reduce nonresponse. The best respondent may differ by module: the household head may answer demographics, while a plot manager answers production and a person responsible for livestock answers animal-health questions. Shorter recall periods or multiple seasonal visits can improve measurement, although they increase cost.

Third, every absence should generate information. Survey teams should record contact attempts, timing, reason for nonresponse, interviewer observations and whether the problem concerned the household, a module or a single item. Such paradata help distinguish refusal from migration, inaccessibility, illness, lack of knowledge or questionnaire burden. FAO operational guidance emphasises monitoring coverage, nonresponse and imputation as core quality indicators (Food and Agriculture Organization [FAO], 2018).

Fourth, analysis must respect the sample design. Survey weights are not optional decorations: they convert sampled units into population estimates and can be adjusted for unequal response using auxiliary information. Standard errors and confidence intervals must also reflect stratification and clustering. The World Bank notes that weighting is required both for unequal inclusion probabilities and to counter undercoverage and nonresponse, while uncertainty estimates are essential for judging reliability (World Bank, n.d.).

Fifth, item missingness should be examined before it is imputed. Analysts should publish missingness rates by region, farm-size class, sex of manager, tenancy status and other relevant groups. Multiple imputation can represent uncertainty better than a single filled value when its assumptions are credible (Little & Rubin, 2019). But no imputation model can recreate a category of farmer that the survey never identified. Coverage repair requires better frames, targeted follow-up, supplementary samples or external data integration.

Finally, results should carry a representation statement: who was in scope, who was excluded by definition, response rates, variables with substantial missingness, adjustment methods and sensitivity of key conclusions. Survey quality should not be judged by response rate alone, because bias depends on how respondents differ from nonrespondents. Nevertheless, rising nonresponse and imputation can weaken household statistics and deserve direct disclosure (Meyer et al., 2015).

Table 3. A practical inclusion checklist for agricultural surveys

Survey stage	Minimum question to ask	Evidence to retain or report
Frame construction	Which cultivators or holdings could be absent from this source?	Frame date, source coverage, deduplication rules and field-verification results.
Listing and selection	Are ownership, cultivation, management and labour being distinguished?	Listing instructions, local terms, eligibility decisions and selection probabilities.
Fieldwork	Who could not be contacted, and why?	Number and timing of attempts, outcome codes, replacements and interviewer paradata.
Data processing	Which edits, deletions and imputations change the analysed population?	Missingness tables, edit logs, imputation flags and distributional checks.
Estimation	Do weights and variance methods match the design and nonresponse process?	Weight construction, calibration variables, effective sample size and confidence intervals.
Dissemination	Would a user know which farmers remain weakly represented?	Scope statement, subgroup response rates, limitations and sensitivity results.

Conclusion: count the absences

Agricultural surveys do more than describe farming; they define the population that policy can see. When missingness is concentrated among informal tenants, remote smallholders, women managers, migrants or farmers with complex livelihoods, the resulting bias is not random noise. It is a systematic narrowing of agricultural reality. The objective is not a perfectly complete dataset—an impossible standard—but an auditable chain of representation. Survey agencies and researchers should design inclusive frames, visit at appropriate times, identify the knowledgeable respondent, record reasons for absence, use design-consistent weighting and imputation, and report sensitivity openly. The most important missing-data question is therefore not “Which value should fill this blank?” It is “Which farmer is missing, why are they missing, and how would our conclusion change if they were visible?”

References

1. Dillon, A., Carletto, G., Gourlay, S., Wollburg, P., & Zezza, A. (2021). Agricultural survey design: Lessons from the LSMS-ISA and beyond. World Bank. https://documents1.worldbank.org/curated/en/570941632979626559/pdf/Agricultural-Survey-Design-Lessons-from-the-LSMS-ISA-and-Beyond.pdf

2. Food and Agriculture Organization of the United Nations. (2018). World programme for the census of agriculture 2020: Volume 2. Operational guidelines (FAO Statistical Development Series No. 16). https://www.fao.org/3/CA1963EN/ca1963en.pdf
3. Gaddis, I., Oseni, G., Palacios-Lopez, A., & Pieters, J. (2021). Measuring farm labor: Survey experimental evidence from Ghana. The World Bank Economic Review, 35(3), 604–634. https://doi.org/10.1093/wber/lhaa012
4. Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). Survey methodology (2nd ed.). Wiley.
5. Little, R. J. A., & Rubin, D. B. (2019). Statistical analysis with missing data (3rd ed.). Wiley. https://doi.org/10.1002/9781119482260
6. Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household surveys in crisis. Journal of Economic Perspectives, 29(4), 199–226. https://doi.org/10.1257/jep.29.4.199
7. Ministry of Agriculture & Farmers Welfare. (2024, December 10). Digital Agriculture Mission. Press Information Bureau, Government of India. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2082787
8. Ministry of Statistics and Programme Implementation. (2021). Situation assessment of agricultural households and land and livestock holdings of households in rural India, 2019 (NSS 77th Round, Report No. 587). Government of India. https://www.mospi.gov.in/national-sample-survey-office
9. World Bank. (n.d.). Sampling, weighting & estimation. Living Standards Measurement Study. Retrieved July 15, 2026, from https://www.worldbank.org/en/programs/lsms/priority-themes/survey-methods/sampling