Livelihood Impact Evaluation of a Mangrove Conservation Support Intervention Along Sorsogon Bay: Evidence on Short-Term Gains, Long-Term Erosion, and Implications for Program Sustainability

Riza L. Garcia[1*], Elmer A. Lorenzana[2], and John Raymond B. Barajas[3]
[*] *riza.garcia@bicol-u.edu.ph*

[1] *Department of Economics*
Bicol University
Legazpi City, Albay, Philippines

[2] *Department of Entrepreneurship*
Bicol University
Legazpi City, Albay, Philippines

[3] *Department of Chemical Engineering*
Bicol University
Legazpi City, Albay, Philippines

Abstract*— Livelihood programs embedded within environmental conservation initiatives are widely implemented across coastal regions in developing countries; however, their long-term effectiveness remains infrequently subjected to rigorous empirical evaluation. In an attempt to address this gap, this study implemented a quasi-experimental impact evaluation of a livelihood support intervention linked to a mangrove reforestation maintenance and monitoring program in coastal communities along Sorsogon Bay, Philippines. To strengthen causal inference, two independent applications of propensity score matching (PSM) were conducted using nearest-neighbor algorithms with data-driven calipers. The first generated 15 matched beneficiary–non-beneficiary pairs for the 2015–2019 comparison, while the second produced 16 matched pairs for the 2015–2024 comparison. Matching was based on twelve baseline covariates encompassing demographic, socioeconomic, and program participation characteristics, thereby constructing a robust counterfactual for impact estimation. Mean peak-season sales—used as a proxy for livelihood performance—were analyzed across three time points: 2015 (baseline), 2019 (midline), and 2024 (endline). Results indicated a statistically significant short-term improvement in peak-season sales among beneficiaries by 2019 (p = 0.01266), demonstrating the initial effectiveness of the intervention. However, this effect was not sustained over time. By 2024, beneficiary sales declined markedly to Php 1,080.00, compared to Php 6,031.30 for the matched comparison group (p = 0.1572), reflecting a reversal of earlier gains and a loss of relative advantage. These findings suggested that while the intervention was capable of generating measurable short-term livelihood improvements, it failed to produce durable economic outcomes. The observed pattern points to underlying structural gaps, particularly in sustained technical support, market integration, and adaptive program design necessary for long-term resilience. Importantly, the evaluation findings were formally presented to the concerned local government unit, which responded constructively by initiating deliberations on program redesign and exploring additional budget allocation. Hence, the present work underscored the role of rigorous impact evaluation not only as a mechanism for accountability, but as a critical input to evidence-informed policy refinement and program strengthening in coastal development contexts.

***Keywords*— impact evaluation, propensity score matching, livelihood programs, mangrove reforestation**

I. Introduction

Coastal ecosystems sustain the livelihoods of millions throughout the developing world, particularly in Southeast Asia, where fishing, aquaculture, and related micro-enterprises form the economic backbone of local communities [1]. Among the most ecologically and economically significant of these ecosystems are mangroves, whose roles extend well beyond carbon sequestration and coastal protection to include direct livelihood provision for households dependent on inshore marine resources [2]. Recognizing this dual function, local government units (LGUs) and conservation organizations across the Philippines have increasingly embedded livelihood support components into mangrove reforestation and maintenance programs — seeking simultaneously to restore degraded ecosystems and to improve the economic well-being of participating communities [3].

Despite the proliferation of such programs, evaluations remain largely restricted in scope and time horizon. Most assessments capture immediate or short-term outcomes — changes in income or productivity measured within one to two years of implementation — without examining whether gains persist once active support has ended [4]. This is a consequential gap. Interventions that initially appear effective may conceal structural weaknesses that lead to the erosion of benefits when programs are scaled back or when communities encounter external shocks such as market disruptions, climate-related events, or ecological change [5]. Without longitudinal evidence, decision-makers lack the information needed to distinguish genuinely transformative programs from those producing only transient improvements [6].

The problem of sustainability is especially pressing for LGU-led programs in the Philippines, where budget constraints, political leadership transitions, and competing development priorities can limit the continuity of intervention support [7]. Evidence-based evaluation that tracks outcomes over a sufficiently long period is therefore essential not only for accountability but for constructive program redesign [6]. Yet the literature on livelihood program sustainability — particularly in the context of community-based conservation — remains thin [8], [9].

In an attempt to address this gap, this study therefore employed a longitudinal, quasi-experimental impact evaluation of a livelihood support intervention implemented in coastal communities along Sorsogon Bay, Philippines. The intervention was embedded within a mangrove reforestation maintenance and monitoring program administered by the local government unit. The primary livelihood outcome examined is peak-season sales — a relevant indicator of enterprise performance in seasonal market contexts. The evaluation spans nine years, covering baseline (2015), midline (2019), and endline (2024) measurements, enabling a rigorous assessment of both short-term and long-term intervention effects.

The contributions of this work are threefold. First, it provides empirical evidence on the temporal dynamics of livelihood intervention effects, demonstrating that early gains do not automatically translate into sustained outcomes. Second, it applies and transparently reports two independent PSM procedures with data-driven calipers in a small-sample, real-world policy evaluation context. Third, it documents how evaluation findings were used to engage the local government in evidence-informed deliberation — a concrete demonstration of impact evaluation as a driver of governance reform rather than merely a reporting exercise.

II. Methodology

A. Participants and Context of the Study

The study was conducted in coastal communities along Sorsogon Bay in the Bicol Region of the Philippines. The bay and its surrounding mangrove zones have long supported the livelihoods of fishing and small-enterprise households, making it a representative setting for evaluating livelihood programs tied to coastal conservation. The livelihood support intervention evaluated here was embedded within an LGU-administered mangrove reforestation, maintenance, and monitoring initiative.

Prior to participant selection and data collection, the necessary institutional and ethical clearances were secured. The research team first established a Memorandum of Understanding with the concerned locality, formalizing the terms of engagement between the research institution and the local government. Participant confidentiality was ensured throughout the process, and the study was assessed to pose negligible risk to all participants. The conduct of the study was subsequently authorized through an Administrative Order issued by Bicol University, after which the research team proceeded with participant engagement and data collection. To further protect participant privacy, all individual identifying details have been anonymized throughout this paper, and all analyses are reported at the aggregate group level.

The full study pool comprised 109 individuals: 40 program beneficiaries (treated group) and 69 non-beneficiaries (control pool). Participants who received livelihood support through the program are referred to as beneficiaries, while community members who were not enrolled serve as the counterfactual group. Both groups shared the same coastal setting and were similarly dependent on seasonally variable income streams tied to local resource availability.

B. Research Design and Rationale for PSM

Because program participation was not randomly assigned, a quasi-experimental design was adopted using propensity score matching (PSM) [10]-[12]. PSM estimates the probability of receiving treatment — here, program participation — as a function of pre-intervention covariates, and then pairs each treated unit with the untreated unit most similar on this estimated probability [11], [13]. By conditioning on a rich set of baseline characteristics, PSM substantially reduces the selection bias that would otherwise confound causal inferences in non-randomized program evaluations [12].

The need for PSM is clearly evident in the pre-matching data. Before any adjustment, the two groups differed substantially on several baseline characteristics. Gender composition showed the largest imbalance, with beneficiaries being far more likely to be female (Std. Mean Difference, SMD = 1.88 across both analyses). Participation in mangrove planting activities also strongly distinguished the groups (SMD $\approx$ 1.05 in the 2019 analysis and 0.99 in the 2024 analysis), as did years of experience in the livelihood activity (SMD $\approx$ -0.87) and age (SMD ≈ -0.69). The overall propensity score distance SMD exceeded 1.94 in both analyses, confirming that raw comparisons between the groups would have been severely biased.

C. Covariates and PSM Specification

Twelve baseline covariates were included in the PSM logistic regression model: age in 2015 (Age_2015), sex, role in the household (Role_in_HH), number of children (Num_Chldrn), years engaged in the livelihood activity (Yrsgln), and four binary indicators for reasons for program participation (resource access, traditional/cultural affinity, economic viability, and scheduling flexibility). Three binary variables capturing participation in mangrove-related activities prior to the program period — mangrove planting (Part_MgrvPlt), clean-up drives (Part_ClnUp), and awareness campaigns (Part_AwrCmpgn) — were also included. Together, these covariates comprehensively characterize the demographic profile, livelihood history, motivational disposition, and prior conservation engagement of each participant.

Nearest-neighbor matching with the "discard both" option was applied, meaning units from either group whose propensity scores fell outside the region of common support were excluded from the analysis prior to matching. This ensured that comparisons are made only among individuals for whom a plausible counterfactual exists, avoiding potentially misleading extrapolations [14], [15].

D. Caliper Specification and Match Quality

The caliper width was set at 0.2 standard deviations of the logit-transformed propensity score — the standard established by Cochran and Rubin's theoretical work on bias reduction through caliper matching [16] and operationalized for the propensity score context by Austin [17] — which has since been widely adopted in applied evaluation research [18]. This was computed as:

$$\varepsilon = 0.2 \times SD(logit(\hat{p}_i))$$

where $\hat{p}_i$ is the estimated propensity score for unit i. Units that could not be matched within this threshold were discarded, prioritizing match quality over sample retention.

Two independent PSM runs were executed, one for each comparison period. The first, used for the 2015–2019 (midline) analysis, yielded 15 matched pairs from the full pool of 109 individuals, with 47 controls and 25 treated units discarded for falling outside common support or beyond the caliper. The second, used for the 2015–2024 (endline) analysis, yielded 16 matched pairs, with 37 controls and 24 treated units discarded. Because the two runs operated independently and applied the caliper to the same underlying pool, the resulting matched samples are not necessarily identical across the two analyses — a transparency consideration discussed further in the limitations section.

Post-matching balance diagnostics confirmed substantial improvements across all covariates in both runs. In the 2019 analysis, the propensity score SMD was reduced from 1.943 to 0.002, and six of twelve covariates achieved perfect balance (SMD = 0.000). Residual imbalances were small: the largest was for years in the livelihood activity (SMD = 0.432), which reflected the boundary conditions imposed by the caliper on a small sample rather than a failure of the matching procedure. In the 2024 analysis, the propensity score SMD dropped from 1.953 to 0.073, with seven of twelve covariates achieving perfect post-match balance. **Table I** presents the pre- and post-match SMDs for selected key covariates across both analyses.

TABLE I. PRE- AND POST-MATCHING STANDARDIZED MEAN DIFFERENCES (SMD) FOR SELECTED COVARIATES

Covariate	Pre-Match SMD (2019)	Post-Match SMD (2019)	Pre-Match SMD (2024)	Post-Match SMD (2024)
Prop. Score	1.943	0.002	1.953	0.073
Age_2015	−0.689	−0.142	−0.689	0.193
Sex	1.882	0.000	1.882	0.400
Role_in_HH	−0.294	0.319	−0.294	0.000
Num_Chldrn	0.061	−0.111	0.061	−0.035
Yrsgln	−0.873	0.432	−0.873	0.417
RSN_Access	−0.391	0.133	−0.391	0.000
RSN_EconViab	0.189	0.000	0.189	0.000
Part_MgrvPlt	1.051	0.000	0.993	0.000
Part_ClnUp	0.775	0.000	0.745	0.000

Note: SMD: Standardized Mean Difference. Values below 0.1 (absolute) are generally considered well-balanced [16]-[18].

E. Outcome Measure and Statistical Tests

The primary outcome variable was mean peak-season sales (SalesPk), denominated in Philippine Pesos (Php). Peak-season sales were selected because they capture the maximum income-generating capacity of micro-enterprise participants and are particularly sensitive to the types of productive and market support that the intervention aimed to provide. Following PSM, differences in mean peak-season sales between beneficiaries and their matched counterfactuals were assessed using paired and independent samples t-tests, applied as appropriate to the comparison structure at each time point. Statistical significance was evaluated against the conventional $\alpha = 0.05$ threshold.

III. RESULTS

A. Baseline Comparability (2015)

At the pre-intervention baseline in 2015, beneficiaries recorded a higher mean peak-season sales figure (Php 13,293.80) compared to their matched counterfactuals (Php 8,776.00). Despite this numerical gap of approximately Php 4,518, the difference was not statistically significant ($p = 0.4878$). This result confirmed that the matched groups were broadly equivalent before the intervention began, validating the PSM procedure and establishing a credible pre-intervention benchmark against which subsequent changes can be assessed.

B. Midline Results (2019) – 15 Matched Pairs

By 2019, four years into the program, a statistically significant divergence had emerged between the two groups based on 15 matched pairs. Beneficiaries recorded mean peak-season sales of Php 12,211.30, compared to Php 8,975.00 for the matched non-beneficiaries, yielding a p-value of 0.01266. This result provided clear evidence of a positive short-term intervention effect. The difference of approximately Php 3,236 represented a meaningful improvement in beneficiaries' earning capacity during peak sales periods. These midline findings constituted early evidence of program effectiveness: the livelihood support components appear to have translated into tangible economic gains in the years following implementation.

C. Endline Results (2024) – 16 Matched Pairs

The endline results, derived from 16 matched pairs, presented a starkly different picture. By 2024, beneficiaries' mean peak-season sales had declined sharply to Php 1,080.00, while the matched counterfactual group maintained considerably higher mean sales of Php 6,031.30. Although the group difference did not cross the conventional significance threshold ($p = 0.1572$), the practical magnitude of the reversal is substantial. Beneficiaries moved from outperforming the comparison group by Php 3,236 at midline to underperforming it by approximately Php 4,951 at endline — a swing of more than Php 8,000 per peak season over a five-year period.

The non-significance of the 2024 difference should not be interpreted as evidence of no effect. With 16 matched pairs and the large variance in sales figures observed at endline, the analysis has limited statistical power to detect effects with precision. The direction and magnitude of the observed difference carry clear policy relevance regardless of the p-value, and are consistent with the interpretation that early intervention gains had fully eroded by the endline period.

D. Trajectory of Effects

Table II and **Fig. 1** illustrates the diverging trajectories of the two groups over the evaluation period. The beneficiary group initially recorded higher sales, maintained a statistically significant advantage at midline, and then experienced a dramatic decline by endline. The counterfactual group, by contrast, exhibited relatively stable sales performance throughout the nine-year window — declining modestly from

Php 8,776.00 to Php 6,031.30 but without the precipitous collapse observed among beneficiaries. This asymmetric pattern suggested that the sharp beneficiary decline reflects forces specific to the beneficiary group rather than a general market deterioration affecting all community members equally.

TABLE II. COMPARISON OF MEAN PEAK-SEASON SALES ACROSS EVALUATION PERIODS

Year	Counterfactual Mean (Php)	Beneficiary Mean (Php)	p-value	Matched Pairs
2015	8,776.00	13,293.80	0.4878	16†
2019	8,975.00	12,211.30	0.01266*	15
2024	6,031.30	1,080.00	0.1572	16

** Statistically significant at α = 0.05. † 2015 baseline comparison uses the 2024 matched sample (16 pairs).*

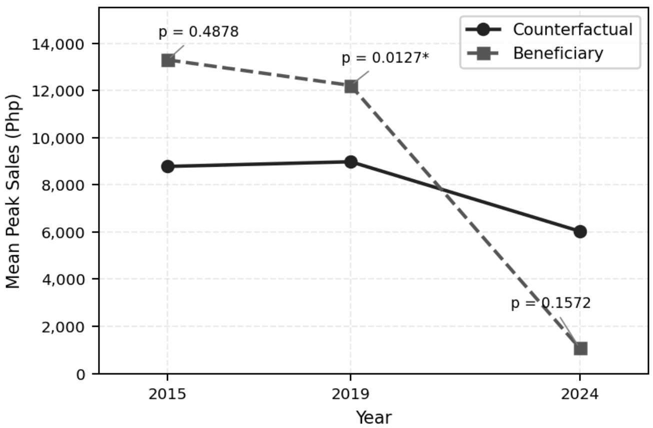

Fig. 1. Mean peak-season sales (Php) of beneficiaries and matched counterfactuals across the three evaluation periods. Asterisk (*) denotes significance at α = 0.05.

IV. DISCUSSION

A. Short-Term Effectiveness

The midline findings confirmed that the intervention was effective in generating short- to medium-term livelihood gains. The statistically significant improvement in beneficiaries' peak-season sales by 2019 is consistent with the broader empirical literature on livelihood support programs, which generally reports positive immediate outcomes attributable to financial assistance, technical training, or improved access to productive resources [9], [19]. The PSM analysis lends additional credibility to this conclusion: given the strong pre-matching imbalance between the groups — particularly on gender composition, conservation activity participation, and years of livelihood experience — the 2019 result would have been far less interpretable without the matching adjustment. Post-matching balance diagnostics further confirmed that the 2019 comparison groups were well-aligned on baseline characteristics, strengthening confidence in the causal interpretation of the midline improvement.

B. Long-Term Sustainability Failure

The dramatic collapse in beneficiaries' sales by 2024 presents a critical challenge to the overall evaluation. Three broad explanations are plausible and likely operate in combination. First, the program may have created reliance on resources or structures that were not internalized as durable productive capacities. When active support was withdrawn, beneficiaries lacked the tools, market connections, or savings buffers needed to maintain their earlier performance levels [5]. This pattern — where supported groups outperform comparison groups during active program phases but subsequently fall below them — is a known risk in short-cycle interventions without explicit sustainability mechanisms [20].

Second, external shocks may have disproportionately affected beneficiaries. The period between 2019 and 2024 coincided with the COVID-19 pandemic, which severely disrupted the livelihoods of coastal and small-enterprise households across the Philippines and contributed to significant income volatility in communities dependent on seasonal markets and local trade [21]. Beneficiaries, potentially more integrated into formal market channels opened by the program, may have faced higher exposure to these disruptions than the less market-integrated comparison group.

Third, the program appeared to have lacked explicit mechanisms for long-term adaptive support. The livelihood literature consistently identifies such mechanisms — including periodic refresher training, revolving credit access, shock-coping safety nets, and links to extension services — as essential components for converting short-term gains into lasting improvements [5], [19]. In their absence, even initially successful interventions are vulnerable to decay over time [5], [20].

C. Policy Engagement and Governance Implications

A distinctive contribution of this evaluation is its documentation of the formal policy engagement process that followed the analysis. The findings were presented to the concerned LGU, which received them not as an indictment but as a basis for institutional learning. This response exemplified the utilization-focused orientation to evaluation advocated in the literature, where findings are designed from the outset to inform and improve rather than merely judge [22], [23]. The LGU's expressed openness to program redesign — including deliberations over additional budget allocation and the integration of adaptive support and resilience-building features — represented precisely the kind of evidence-to-policy translation that evaluation frameworks aspire to produce [24].

This outcome is also instructive from a governance standpoint. It illustrated that unfavorable evaluation results — including evidence of long-term program failure — can be productively absorbed when the evaluation process is participatory, transparent, and explicitly oriented toward improvement. The positive receptance of the findings by the LGU suggested that a culture of evidence-informed decision-making is being cultivated in this governance context, which bodes well for the quality of future interventions in the area [25].

V. LIMITATIONS AND FUTURE WORK

While this study employed a rigorous quasi-experimental design and takes deliberate steps to ensure methodological transparency, several important limitations must be acknowledged.

- **Small matched sample and limited statistical power**: The most consequential limitation of this study is the small number of matched pairs — 15 in the 2019 analysis and 16 in the 2024 analysis. Although these sample sizes reflected the targeted and localized nature of the intervention, they substantially limit statistical power. This was most evident at the endline, where the observed difference in peak-season sales between beneficiaries and the counterfactual group (Php 4,951) is practically large but does not reach conventional significance ($p = 0.1572$). With as few as 15–16 pairs, the minimum detectable effect at 80% power and $\alpha = 0.05$ is considerably larger than the differences likely to be detected in a more adequately powered study. Future evaluations of similar programs should plan for larger enrollment and matched sample targets to ensure adequate power across all time points.

- **Non-identical matched samples across time points**: Because two independent PSM runs were conducted — one for each comparison period — the matched pairs used in the 2019 and 2024 analyses are not guaranteed to be identical. The two runs produced somewhat different matched subsets from the same underlying pool, with slightly different numbers of discarded units. This meant that the temporal comparison of midline and endline effects cannot be treated as a fully longitudinal panel, and some of the observed change over time may partly reflect differences in sample composition between analyses rather than purely temporal dynamics. Panel-based approaches in future studies would resolve this limitation.

- **Quasi-complete separation in propensity score estimation**: Both PSM runs triggered the warning "glm.fit: fitted probabilities numerically 0 or 1 occurred," indicating quasi-complete separation in the logistic regression used to estimate propensity scores. This condition arises when one or more covariate patterns — in this case, participation in mangrove-related activities such as planting (Part_MgrvPlt), clean-up drives (Part_ClnUp), and awareness campaigns (Part_AwrCmpgn), for which control units consistently recorded zero — perfectly or near-perfectly predict treatment assignment. Quasi-complete separation can produce propensity score estimates that are numerically unstable for affected units. The application of the caliper and the discard of out-of-support units mitigated but did not fully eliminate this concern, as the propensity scores used in matching for retained units could still be influenced by the extreme patterns in the estimation sample. Users of these results should interpret the matched comparisons with appropriate caution, particularly for subgroups with extreme covariate patterns.

- **Narrow outcome measure**: Peak-season sales, while a relevant and tractable indicator of enterprise performance, represented a narrow slice of livelihood well-being. Comprehensive assessments should incorporate income diversity, household consumption, savings, food security, and vulnerability indicators to capture the full range of intervention effects. The absence of lean-season sales analysis, despite such data being available in the study dataset (SalesLn), also meant that the evaluation captures only one dimension of the seasonal income cycle. Future work should examine whether the patterns observed for peak-season sales are replicated or diverge for lean-season income.

- **Absence of qualitative data**: The study relied entirely on quantitative sales records and does not incorporate qualitative evidence on the mechanisms underlying the observed outcomes. The explanations offered for the endline decline — program dependency, external shocks, and lack of sustainability mechanisms — are theoretically grounded and empirically consistent but remain inferential. Qualitative follow-up with beneficiaries and program administrators would help to identify which factors were most operative and would provide more actionable guidance for program redesign.

- **Unobserved confounding**: PSM eliminates selection bias only on observable characteristics. Unobserved factors that simultaneously influenced both program participation and peak-season sales performance — such as individual entrepreneurial motivation, household wealth, or informal network ties — may still confound the estimated effects. While the rich set of 12 baseline covariates used in the matching model helped to capture many plausible confounders, residual unobserved heterogeneity cannot be ruled out. This is an inherent limitation of all observational quasi-experimental designs and should be borne in mind when interpreting the causal claims made in this paper.

VI. Conclusion

This study evaluated the short-term and long-term livelihood impacts of a mangrove conservation-linked support intervention implemented in coastal communities along Sorsogon Bay, Philippines. Two independent applications of propensity score matching with data-driven calipers — one yielding 15 matched pairs for the midline analysis and another yielding 16 pairs for the endline analysis — were used to construct statistically comparable counterfactual groups from a pool exhibiting severe pre-matching imbalance. Across twelve baseline covariates including gender, age, livelihood experience, participation motivation, and conservation activity history, post-matching balance was substantially improved in both analyses, lending credibility to the subsequent outcome comparisons.

The results revealed a clear and significant positive effect at the midline period in 2019 ($p = 0.01266$), followed by a pronounced and practically significant reversal by the endline in 2024. Beneficiaries, who had outperformed their matched counterparts during the active program phase, recorded dramatically lower sales nine years after baseline — a finding that underscored the structural fragility of intervention gains in the absence of long-term support, adaptive mechanisms, and shock resilience. The non-significance of the 2024 difference ($p = 0.1572$) reflected primarily the limited statistical power of the small matched sample and should not be misconstrued as evidence of no effect.

The study also demonstrated the value of translating rigorous evaluation findings into direct policy engagement. By presenting results to the concerned local government unit and receiving a constructive response oriented toward program redesign rather than discontinuation, the evaluation fulfilled its most important purpose: generating actionable evidence for the people responsible for governing the communities it examined. Future livelihood programs embedded within environmental conservation initiatives should design explicitly for long-term sustainability from the outset — building adaptive capacities, market linkages, and shock resilience as integral components rather than afterthoughts.

ACKNOWLEDGMENT

The author wishes to express sincere gratitude to Bicol University, the Research Development and Management Division, the Center for Policy Studies and Development, the Department of Economics, the Department of Chemical Engineering, and the Department of Entrepreneurship for their institutional support in the conduct of this research. Special thanks are due to Dr. Lany L. Maceda, Vice-President for Research, Development and Extension, and Dr. Baby Boy Benjamin D. Nebres III, SUC President IV, for their guidance and encouragement throughout the study.

REFERENCES

[1] M. Fabinyi et al., "Coastal transitions: Small-scale fisheries, livelihoods, and maritime zone developments in Southeast Asia," Journal of Rural Studies, vol. 91, pp. 184–194, Apr. 2022, doi: 10.1016/j.jrurstud.2022.02.006.

[2] H. Lee, H. Kim, E. Park, and B. Lee, "Beyond carbon: a systematic review of multiple ecosystem services of mangroves," J Coast Conserv, vol. 29, no. 6, Oct. 2025, doi: 10.1007/s11852-025-01148-4.

[3] R. B. Valenzuela, Y. Yeo-Chang, M. S. Park, and J.-N. Chun, "Local People's Participation in Mangrove Restoration Projects and Impacts on Social Capital and Livelihood: A Case Study in the Philippines," Forests, vol. 11, no. 5, p. 580, May 2020, doi: 10.3390/f11050580.

[4] J. Brooks, K. A. Waylen, and M. B. Mulder, "Assessing community-based conservation projects: A systematic review and multilevel analysis of attitudinal, behavioral, ecological, and economic outcomes," Environ Evid, vol. 2, no. 1, Jan. 2013, doi: 10.1186/2047-2382-2-2.

[5] P. Daly et al., "Rethinking relief, reconstruction and development: Evaluating the effectiveness and sustainability of post-disaster livelihood aid," International Journal of Disaster Risk Reduction, vol. 49, p. 101650, Oct. 2020, doi: 10.1016/j.ijdrr.2020.101650.

[6] E. Woodhouse et al., "Guiding principles for evaluating the impacts of conservation interventions on human well-being," Phil. Trans. R. Soc. B, vol. 370, no. 1681, p. 20150103, Nov. 2015, doi: 10.1098/rstb.2015.0103.

[7] R. S. Pomeroy, E. G. Oracion, R. B. Pollnac, and D. A. Caballes, "Perceived economic factors influencing the sustainability of integrated coastal management projects in the Philippines," Ocean & Coastal Management, vol. 48, no. 3–6, pp. 360–377, 2005, doi: 10.1016/j.ocecoaman.2005.04.010.

[8] D. Roe et al., "Are alternative livelihood projects effective at reducing local threats to specified elements of biodiversity and/or improving or maintaining the conservation status of those elements?," Environ Evid, vol. 4, no. 1, Nov. 2015, doi: 10.1186/s13750-015-0048-1.

[9] B. Fariss, N. DeMello, K. A. Powlen, C. E. Latimer, Y. Masuda, and C. M. Kennedy, "Catalyzing success in community‐based conservation," Conservation Biology, vol. 37, no. 1, Aug. 2022, doi: 10.1111/cobi.13973.

[10] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," Biometrika, vol. 70, no. 1, pp. 41–55, 1983, doi: 10.1093/biomet/70.1.41.

[11] M. Caliendo and S. Kopeinig, "Some Practical Guidance For The Implementation Of Propensity Score Matching," Journal of Economic Surveys, vol. 22, no. 1, pp. 31–72, Jan. 2008, doi: 10.1111/j.1467-6419.2007.00527.x.

[12] S. Guo, M. Fraser, and Q. Chen, "Propensity Score Analysis: Recent Debate and Discussion," Journal of the Society for Social Work and Research, vol. 11, no. 3, pp. 463–482, Sep. 2020, doi: 10.1086/711393.

[13] B. Y. Andrew, M. Alan Brookhart, R. Pearse, K. Raghunathan, and V. Krishnamoorthy, "Propensity score methods in observational research: brief review and guide for authors," British Journal of Anaesthesia, vol. 131, no. 5, pp. 805–809, Nov. 2023, doi: 10.1016/j.bja.2023.06.054.

[14] E. A. Stuart, "Matching Methods for Causal Inference: A Review and a Look Forward," Statist. Sci., vol. 25, no. 1, Feb. 2010, doi: 10.1214/09-sts313.

[15] D. E. Ho, K. Imai, G. King, and E. A. Stuart, "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference," J. Stat. Soft., vol. 42, no. 8, 2011, doi: 10.18637/jss.v042.i08.

[16] W. G. Cochran and D. B. Rubin, "Controlling bias in observational studies: A review," Sankhyā: The Indian Journal of Statistics, Series A, vol. 35, no. 4, pp. 417–446, 1973. [Online]. Available: http://www.jstor.org/stable/25049893

[17] P. C. Austin, "Optimal caliper widths for propensity‐score matching when estimating differences in means and differences in proportions in observational studies," Pharmaceutical Statistics, vol. 10, no. 2, pp. 150‐161, Mar. 2011, doi: 10.1002/pst.433.

[18] M. Lunt, "Selecting an Appropriate Caliper Can Be Essential for Achieving Good Balance With Propensity Score Matching," American Journal of Epidemiology, vol. 179, no. 2, pp. 226–235, Oct. 2013, doi: 10.1093/aje/kwt212.

[19] A. Banerjee et al., "A multifaceted program causes lasting progress for the very poor: Evidence from six countries," Science, vol. 348, no. 6236, May 2015, doi: 10.1126/science.1260799.

[20] D. H. Bailey, G. J. Duncan, F. Cunha, B. R. Foorman, and D. S. Yeager, "Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions," Psychol Sci Public Interest, vol. 21, no. 2, pp. 55–97, Oct. 2020, doi: 10.1177/1529100620915848.

[21] R. J. Maliao, P. R. Fernandez Jr, and R. F. Subade, "Well-being of artisanal fishing communities and children's engagement in fisheries amidst the COVID-19 pandemic: a case in Aklan, Philippines," Humanit Soc Sci Commun, vol. 10, no. 1, May 2023, doi: 10.1057/s41599-023-01716-9.

[22] M. Q. Patton, "Utilization-Focused Evaluation," International Handbook of Educational Evaluation. Springer Netherlands, pp. 223–242, 2003. doi: 10.1007/978-94-010-0309-4_15.

[23] M. Q. Patton, Utilization-focused evaluation, 4th ed. Thousand Oaks, CA, USA: Sage Publications, 2008.

[24] S. Bowen and A. B. Zwi, "Pathways to 'Evidence-Informed' Policy and Practice: A Framework for Action," PLoS Med, vol. 2, no. 7, p. e166, May 2005, doi: 10.1371/journal.pmed.0020166.

[25] P. Cairney and K. Oliver, "How Should Academics Engage in Policymaking to Achieve Impact?," Political Studies Review, vol. 18, no. 2, pp. 228–244, Nov. 2018, doi: 10.1177/1478929918807714.