Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Household Satisfaction and Compensation Preferences under the Grain-for-Green Program: Evidence from Guyuan on China's Loess Plateau

miao wu
Ningxia Academy of Agriculture and Forestry Sciences
zhijia yu
Ningxia Academy of Agriculture and Forestry Sciences
jie ma
Ningxia Academy of Agriculture and Forestry Sciences
zhengan wang
wumiao25@mails.ucas.ac.cn

Ningxia Academy of Agriculture and Forestry Sciences

Research Article

Keywords: Grain-for-Green Program, Payments for ecosystem services, Compensation dissatisfaction, Policy familiarity, Maintenance arrangements, Loess Plateau

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11550116/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Household evaluations provide important evidence for diagnosing problems in ecological compensation programs. However, concerns about payment standards need to be distinguished from overall dissatisfaction before policy responses are designed. Taking Guyuan on China's Loess Plateau as a case study, this research used a 2024 convenience survey of 1,650 Grain-for-Green households to examine compensation evaluations and preferences. Descriptive analysis and multivariable Logit models (n = 1,601 complete cases) were combined with locality-clustered uncertainty estimation, multiple-comparison correction and sensitivity analyses. The results show that (1) 39.7% of surveyed households were satisfied, 32.4% were neutral and 27.9% were dissatisfied, with substantial differences across sampled counties and districts. (2) Low payment standards were reported by 1,588 households (96.2%); nevertheless, 71.6% of this group were satisfied or neutral overall. Concern about payment amounts therefore did not necessarily indicate overall dissatisfaction. (3) Jingyuan households had higher odds of dissatisfaction than Longde households (OR = 23.31), while respondents who were not very familiar with policy had higher odds than those who were very familiar (OR = 11.82). Both contrasts retained statistical support after multiple-comparison correction and exclusion of Pengyang, and the familiarity contrast persisted in locality fixed-effect models. These estimates did not establish a monotonic familiarity gradient. Evidence for individual maintenance depended on model specification and multiplicity correction, while total income category showed no clear common linear association. (4) Among households' preferred compensation arrangements, cash payments were the most frequently selected option (86.0%); employment opportunities (28.5%) and vocational training (19.9%) were also preferred. This study provides empirical evidence for distinguishing payment-specific concerns from overall evaluations and identifying priorities for local investigation before compensation arrangements are redesigned.

1. Introduction

Forest restoration can generate environmental benefits that extend beyond the landholders who bear its costs. Payments for ecosystem services (PES) seek to connect these benefits with incentives for land management, making both environmental effectiveness and the distribution of costs and benefits central to programme design (Engel et al. 2008). China's Grain-for-Green Program, also known as the Sloping Land Conversion Program, uses public payments to engage rural households in cropland conversion and restoration(Bennett et al. 2008). Such programmes are implemented through local arrangements that can differ in consultation, monitoring and support. For example, Bennett et al. (2014) found associations between household-reported tree survival, available labour and local implementation arrangements. This evidence suggests that compensation assessment should consider the settings in which payments are delivered alongside the amounts offered. Household evaluations provide another perspective on these arrangements by recording participants' assessments of the compensation they receive. Examining variation in these evaluations can help identify issues requiring closer investigation, while remaining distinct from assessments of ecological performance.

Monetary valuation and evaluations of existing compensation address different questions. Lindhjem and Mitani (2012) used contingent valuation to examine Norwegian forest owners' willingness to accept compensation for voluntary conservation, while Wang et al. (2026) elicited residents' willingness to accept or pay for multifunctional forest compensation in the Jinghe River Basin. These studies concern stated monetary preferences under specified conservation scenarios. An assessment that existing payment standards are low is instead a judgement about payment levels, while an overall satisfaction item requests a general evaluation without specifying which aspects respondents should weigh. Treating these responses as interchangeable could obscure differences in the meaning attached to compensation. In Costa Rica, Chapman et al.(2020) found that farmer participants often understood PES payments as support for land stewardship rather than solely as transactions for environmental services. A survey experiment by Clot et al. (2017) also found that the terms used to describe a proposed PES programme could influence its favourability ratings. These findings make participants' interpretations relevant to the study of compensation evaluations, but do not establish why payment concerns and overall satisfaction might diverge in a particular setting. The present study therefore examines the two responses separately and asks whether households sharing a concern about low payment standards nevertheless report different overall evaluations. This provides a basis for identifying issues requiring follow-up, without treating either response alone as a measure of an objective compensation shortfall.

Household income and policy familiarity offer complementary ways of examining variation in compensation evaluations, but require careful interpretation. Li et al. (2011) reported that income gains under the Sloping Land Conversion Program did not coincide with increased labour transfer towards nonfarm activities at their study site. Uchida et al. (2009) found that off-farm labour responses varied with initial human and physical capital, whereas Kelly and Huo (2013) attributed increased nonfarm employment to labour substitution rather than relaxation of constraints. These findings motivate attention to livelihood structure alongside total income. Households in the same income category may differ in agricultural dependence, labour commitments and access to alternative employment, providing a testable basis for investigating different compensation evaluations. Policy familiarity raises a related measurement issue. Household preferences can depend on the conditions of PES contracts(Li et al. 2017), while equity research highlights the importance of assessing the distribution of benefits and burdens(Pascual et al. 2014). Self-reported familiarity captures respondents' perceived understanding; it does not independently demonstrate knowledge of entitlements or agreement with programme rules. A well-informed household may still object to payment standards or procedures. Familiarity may also reflect contact with implementing organisations or prior programme experience. These possibilities justify examining its association with dissatisfaction without assuming that better information necessarily produces greater approval. These household-level questions also need to be situated within the local arrangements governing policy communication, payment delivery and land maintenance.

Local implementation provides a further reason to examine compensation evaluations within their geographical and organisational settings. He and Lang (2015) identified gaps between programme objectives and implementation in Southwest China, highlighting problems in land and household

selection and limitations in local participation. He (2021) examined how formal programme arrangements interacted with pre-existing local institutions, while Zinda and Zhang (2019) documented how governance, livelihoods and environmental conditions combined to shape tree-cover change. These studies address implementation, institutions or vegetation outcomes rather than the compensation evaluations examined here. Nevertheless, they identify dimensions worth investigating: who communicates programme rules, how payments are delivered, and who undertakes land maintenance. Households under different maintenance arrangements may encounter different responsibilities and forms of support, but the arrangement alone does not measure their actual workload or expenditure. Geographical differences in evaluations may likewise reflect several aspects of implementation alongside household and environmental conditions. Comparing evaluations across maintenance arrangements and places can therefore identify where closer investigation is warranted, while differences between places alone do not establish their causes.

Against this background, this study uses a 2024 cross-sectional survey of 1,650 Grain-for-Green households in Guyuan, Ningxia, on the Loess Plateau, to examine compensation evaluations as evidence for local policy diagnosis. It addresses three research questions: (1) How do overall compensation evaluations vary among households reporting low payment standards? (2) How is explicit dissatisfaction associated with maintenance arrangements, self-reported policy familiarity and location in models including the measured household characteristics? (3) Which associations retain statistical support across alternative samples and geographical specifications when multiple comparisons are taken into account? Descriptive cross-tabulations examine evaluations within the low-payment-standard subgroup, while multivariable models examine dissatisfaction among complete cases from the full survey. Descriptive analysis also summarises preferences for compensation arrangements. Sensitivity analyses assess dependence on the inclusion of Pengyang and on the use of county indicators or recorded-locality fixed effects, with multiplicity correction applied to the reported model families. The analysis is exploratory. Its empirical contribution is to distinguish payment-specific concerns from overall evaluations and assess the consistency of associations that can guide further investigation. The aim is to provide an evidence base for identifying problems requiring verification before compensation arrangements are redesigned.

2. Materials and Methods

2.1. The Study Area

The study was conducted in Guyuan City, southern Ningxia Hui Autonomous Region, northwestern China. Located in the Liupan Mountain region of the Loess Plateau, Guyuan extends from 35°14' to 36°31' N and from 105°19' to 106°57' E (General Office of Guyuan Municipal People's Government 2024). Its landscape comprises loess hills, dissected gullies, mountains and river valleys. The region has a warm-temperate continental monsoon climate, with mean annual temperature, precipitation and evaporation of approximately 7°C, 499.7 mm and 1,364.4 mm, respectively. Precipitation varies considerably across the

municipality and generally decreases from south to north(General Office of Guyuan Municipal People's Government 2024).

Guyuan has implemented several ecological restoration initiatives, including Grain for Green, natural forest protection and watershed rehabilitation. According to official statistics for the end of 2020, land designated for forestry covered approximately 411,073 ha, of which 256,940 ha was forest, with a reported forest coverage rate of 24.42% (General Office of Guyuan Municipal People's Government 2022). The municipality had a permanent resident population of 1,151,900 at the end of 2022(General Office of Guyuan Municipal People's Government 2023). In 2024, its gross domestic product reached CNY 46.373 billion, representing annual growth of 6.1% at constant prices, while annual per capita disposable income among rural residents was CNY 16,612(Guyuan Municipal Bureau of Statistics 2025).

The household survey covered Longde County, Jingyuan County, Pengyang County and Yuanzhou District. All surveyed households had participated in the Grain-for-Green Program. The study focused on their evaluations of compensation and the variation in these evaluations across surveyed locations, maintenance arrangements and household characteristics.

2.2. Questionnaire Design and Survey Procedures

2.2.1. Questionnaire Design

A structured questionnaire was used to collect information on household characteristics, participation in the Grain-for-Green Program, compensation evaluations and policy familiarity. The present analysis focused on overall compensation satisfaction and its associations with maintenance arrangements, policy familiarity, geographical location and household characteristics. The questionnaire covered three main domains:

(1) Household demographic and socioeconomic characteristics. Questions recorded respondents' sex, age, educational attainment and occupation, together with household size, annual household income and income sources. Annual household income was recorded in five categories: below CNY 20,000, CNY 20,000–30,000, CNY 30,000–50,000, CNY 50,000–80,000 and above CNY 80,000. Educational attainment ranged from no formal education to postgraduate education. County, township and village information was also recorded to identify the geographical location of each household.

(2) Program participation, maintenance arrangements and compensation evaluations. Questions covered land enrolled in the program, planted tree species, maintenance arrangements and income from forest products. Maintenance arrangements were classified as forest-farm maintenance, township or village collective maintenance, individual maintenance, or other arrangements. Overall compensation satisfaction was measured using a five-category item ranging from very satisfied to very dissatisfied, with a middle response category retained separately. A subsequent question listed concerns about compensation: low payment standards, delayed payments, limitations on program duration, limited perceived program outcomes, and other reasons. Although labelled as a question about reasons for

dissatisfaction, it was available to all respondents, irrespective of their preceding satisfaction response. Respondents could leave it blank or indicate that they had no dissatisfaction; both were represented by blank entries in the coding scheme. Consequently, a recorded concern about payment standards was considered separately from overall compensation satisfaction.

(3) Policy familiarity and perceptions of compensation arrangements. Respondents rated their familiarity with the relevant policy using four categories: very familiar, basically familiar, not very familiar and unfamiliar. Additional questions addressed information sources, perceived problems with compensation arrangements, preferred forms of support and perceived forest ecosystem functions. The latter items included multiple-response questions. Policy familiarity was measured as a self-assessment rather than an objective knowledge test. In the present analysis, familiarity and overall compensation satisfaction were retained as separate measures, while variable coding and their use in the statistical models are described below.

2.2.2. Sampling and Survey Procedures

A cross-sectional household survey was conducted approximately between May and October 2024 in Longde County, Jingyuan County, Pengyang County and Yuanzhou District. Villages were selected through the research team's existing contacts with village Party branch secretaries. Within these villages, recruitment sought to reach households participating in the Grain-for-Green Program that could be contacted. This constituted convenience sampling: households were not selected from a complete sampling frame, and their inclusion probabilities were unknown.

All surveyed households had participated in the program, and one questionnaire was completed per household. Respondents were informed of the purpose of the study and participated voluntarily. Questionnaires were completed by respondents themselves, with village cadres present during completion. These arrangements facilitated access to participating households, although recruitment through local contacts and the presence of cadres may have influenced both sample composition and respondents' evaluations.

Following data preparation, the dataset used for descriptive analysis comprised 1,650 household records: 829 from Longde, 337 from Jingyuan, 188 from Pengyang and 296 from Yuanzhou. These numbers represent the households retained for analysis rather than county-level sampling quotas. The sample provides evidence on compensation evaluations among the surveyed participating households; it was not designed to produce statistically representative estimates for all participating households in Guyuan.

2.3. Data Preparation and Statistical Analysis

2.3.1. Data Quality Control and Analytical Sample

The descriptive analysis included 1,650 households with recorded compensation-satisfaction responses. Overall satisfaction and reported concerns about compensation were examined as separate measures.

Accordingly, households reporting low payment standards while expressing satisfaction or neutrality overall were retained, as this combination did not necessarily represent an inconsistent response.

Regression analyses used complete cases for all variables included in the models. One household reporting “other” maintenance was excluded from regression because this category was too small for a separate comparison, and 48 additional households were excluded because age information was missing. The main regression sample therefore comprised 1,601 households. Missing covariates were not statistically imputed. Applying the same criteria after excluding Pengyang yielded 1,419 households for the sensitivity analysis. Households excluded from regression remained in the descriptive analysis.

Categorical variables were summarized as frequencies and percentages, and age as the mean, standard deviation and range. Overall compensation evaluations were grouped as satisfied, neutral or dissatisfied by combining the two satisfied categories and the two dissatisfied categories while retaining the middle category separately. Cross-tabulations described evaluations by county-level jurisdiction and, within the subgroup reporting low payment standards, by maintenance arrangement and policy familiarity. Percentages used the household count in the relevant county or subgroup; age-group percentages used respondents with recorded age. For multiple-response questions on income sources, policy information sources, compensation funding sources and preferred compensation arrangements, percentages used all households in the corresponding overall or county sample, rather than the total number of selections. Records with no selected option were reported separately. The reported cross-tabulations were descriptive and did not include hypothesis tests.

2.3.2. Econometric Model Specification

A binary logistic regression model was used to examine associations between household characteristics and overall compensation dissatisfaction. The dependent variable, Y_i, was coded as 1 when respondent i selected either dissatisfied response category and as 0 when the response was satisfied or neutral. This specification distinguished explicit dissatisfaction from the remaining evaluations; it did not classify neutral responses as satisfaction. The three-category distribution of satisfied, neutral and dissatisfied responses was retained for descriptive analysis.

The model was specified as:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \varvec{\beta}_C^{\mathrm{T}}\mathbf{C}_i + \varvec{\beta}_M^{\mathrm{T}}\mathbf{M}_i + \varvec{\beta}_F^{\mathrm{T}}\mathbf{F}_i + \varvec{\gamma}^{\mathrm{T}}\mathbf{X}_i$$

1

where $p_i = \Pr\left(Y_i = 1 \mid \mathrm{C}_i, \mathrm{M}_i, \mathrm{F}_i, \mathrm{X}_i\right)$ is the conditional probability of compensation dissatisfaction; β_0 is the intercept; and the remaining coefficients describe associations with county, maintenance arrangements, policy familiarity and household characteristics.

The vector C_i contained county indicators, with Longde as the reference. The vector M_i represented forest-farm and individual maintenance, with township or village collective maintenance as the reference. Policy familiarity, F_i, was entered using indicators for basically familiar, not very familiar and unfamiliar respondents, with very familiar respondents as the reference. This categorical specification allowed separate familiarity contrasts without imposing a linear or monotonic relationship.

The household covariate vector, X_i, included annual household income category, educational attainment, age, sex, receipt of forest-product income and household size. Income and education were entered as ascending category scores, coded 1–5 and 1–7, respectively. Their coefficients therefore represented a common change in log odds per category increase, rather than a change per monetary unit or year of schooling. Age was expressed in decades and household size as the number of household members. Female respondents and households without forest-product income were the reference groups for the corresponding binary variables. All explanatory variables were entered simultaneously, and the reported specification contained main effects without interaction terms.

The model was estimated by maximum likelihood. Coefficients were exponentiated to obtain odds ratios:

$$\mathrm{OR}_j = \exp\left(\beta_j\right)$$

2

For indicator variables, an odds ratio compared the specified category with its reference group, holding the other included variables constant. For scored or numerical variables, it represented the association per unit increase on the stated scale. Odds ratios were interpreted as associations from the multivariable model rather than probability ratios or causal effects.

Model checks assessed coefficient identifiability, convergence and complete or quasi-complete separation. Uncertainty estimates used HC1 cluster-robust covariance matrices with a cluster-count correction, allowing for within-locality dependence across 30 recorded locality groups defined jointly by county and standardized locality labels. Two-sided tests and 95% confidence intervals used a t distribution with $G - 1$ degrees of freedom, where G was the number of locality groups.

Analyses were conducted in R version 4.5.2. Models were fitted using the stats package; clustered covariance matrices were calculated using sandwich (Zeileis et al. 2020), separation was checked using detectseparation and figures were produced using ggplot2.

2.3.3. Sensitivity Analyses and Multiplicity Control

Two sensitivity analyses examined the dependence of the reported associations on sample composition and geographical specification. First, the Logit model was repeated after excluding Pengyang, which was represented by a single recorded locality, while retaining the same variable definitions and complete-case criteria. Second, linear probability models were estimated by ordinary least squares, replacing

county indicators with recorded-locality fixed effects while retaining the maintenance, familiarity and household covariates. These models were estimated in both the full regression sample and the sample without Pengyang. Coefficients were multiplied by 100 and reported as differences in the probability of dissatisfaction in percentage points. Both sensitivity analyses used the uncertainty-estimation procedure described in Section 2.3.2, with 30 locality groups in the full regression sample and 29 after excluding Pengyang.

To account for multiple comparisons, the Benjamini–Hochberg procedure(Benjamini and Hochberg 1995) was applied jointly to the P values for the 27 non-intercept coefficients across the full-sample Logit model and the model without Pengyang. A separate correction family comprised the P values for the 22 maintenance, familiarity and household coefficients across the two linear probability models with recorded-locality fixed effects; locality-indicator coefficients and intercepts were excluded. Nominal two-sided P values and 95% confidence intervals were reported alongside BH q values, with $q < 0.05$ used to identify comparisons meeting the multiplicity threshold. These corrections covered the two reported model families and did not cover earlier exploratory analyses. The analyses were exploratory and did not test preregistered directional hypotheses.

Complete coefficients from the main and sensitivity models are provided in Supplementary Table S1.

3. Results

3.1. Basic Characteristics of the Sample

The sample comprised 1,650 households participating in the Grain-for-Green Program: 829 in Longde, 296 in Yuanzhou, 337 in Jingyuan and 188 in Pengyang (Table 1).

Table 1
Geographical distribution of surveyed households and regression samples.

Jurisdiction	Surveyed households	Recorded localities	Regression sample
	n	n	Model n
Longde	829	10	816
Yuanzhou	296	16	289
Jingyuan	337	3	314
Pengyang	188	1	182
Total	1650	30	1601

Notes: Model n denotes complete cases after excluding one household reporting other maintenance and 48 households with missing age. Recorded localities are groups defined jointly by county and standardized locality labels; they have not been independently verified as administrative villages. Counts describe the surveyed households rather than county populations. Yuanzhou is a district.

Respondents were predominantly male (1,314; 79.6%), with the male proportion ranging from 76.2% in Longde to 96.3% in Pengyang. Age information was available for 1,602 respondents, whose mean age was 51.6 years (SD = 9.8; range: 20–82 years). The largest age group was 50–59 years (640 respondents; 40.0%), followed by 40–49 years (469; 29.3%). Together, these two groups accounted for 69.2% of respondents with recorded age information (Supplementary Table S2, panel A).

Educational attainment was concentrated at the junior secondary level or below. Primary and junior secondary education were reported by 535 (32.4%) and 530 (32.1%) respondents, respectively, while 265 (16.1%) had no formal education. Together, these categories accounted for 80.6% of the sample. A further 200 respondents (12.1%) reported high-school education, and 120 (7.3%) reported college-level education or above. Regarding occupation, 1,596 respondents (96.7%) selected the farmer category, while the remaining 54 (3.3%) were distributed across other occupational categories (Supplementary Table S2, panel A).

Annual household income most commonly fell within the CNY 20,000–30,000 category (692 households; 41.9%), followed by CNY 30,000–50,000 (453; 27.5%). A further 233 households (14.1%) reported income below CNY 20,000, while 241 (14.6%) and 31 (1.9%) reported CNY 50,000–80,000 and above CNY 80,000, respectively. In the multiple-response question on household income sources, crop cultivation and livestock farming were selected by 852 households (51.6%), temporary employment by 665 (40.3%), regular salaried employment by 121 (7.3%), and other sources by 56 (3.4%). No income-source option was recorded for 217 households (13.2%). All income-source percentages were calculated using the full sample of 1,650 households, and respondents could select more than one option (Supplementary Table S2, panel B).

3.2. Policy Familiarity and Information Sources

Most respondents described themselves as basically familiar with the relevant policy (1,347; 81.6%), while 121 (7.3%) reported being very familiar. A further 158 respondents (9.6%) were not very familiar, and 24 (1.5%) were unfamiliar. Overall, 1,468 respondents (89.0%) selected either the very familiar or basically familiar category, with the latter accounting for the predominant share of responses (Supplementary Table S3, panel A; Figure S1A).

The basically familiar category predominated in all four sampled jurisdictions, accounting for 77.8% of respondents in Longde, 86.1% in Yuanzhou and 76.9% in Jingyuan. All 188 respondents in Pengyang selected this category. Longde had the highest proportion of very familiar respondents (102; 12.3%), compared with 10 (3.4%) in Yuanzhou and nine (2.7%) in Jingyuan. Conversely, Jingyuan had the highest proportions of respondents reporting that they were not very familiar (51; 15.1%) or unfamiliar (18; 5.3%). The corresponding proportions were 9.2% and 0.7% in Longde, and 10.5% and 0.0% in Yuanzhou, respectively (Supplementary Table S3, panel A).

Government publicity was the most frequently reported source of policy information (1,399 respondents; 84.8%), followed by television and radio (290; 17.6%) and the internet (98; 5.9%). Personal or family

experience was selected by 10 respondents (0.6%), and other sources by three (0.2%). Respondents could select multiple sources, and all percentages were calculated using the full sample of 1,650 households rather than the total number of selections. No information-source option was recorded for 120 households (7.3%) (Supplementary Table S3, panel B; Figure S1B).

3.3. Household Evaluations of Compensation

3.3.1. Overall Evaluations and Reported Compensation Concerns

Overall, 655 of the 1,650 surveyed households (39.7%) were satisfied with compensation, 535 (32.4%) were neutral and 460 (27.9%) were dissatisfied. The distribution of evaluations varied across the sampled county-level jurisdictions (Fig. 1). Dissatisfaction was reported by 69 of 829 households in Longde (8.3%) and 49 of 296 in Yuanzhou (16.6%), compared with 222 of 337 in Jingyuan (65.9%) and 120 of 188 in Pengyang (63.8%). No surveyed household in Pengyang selected either satisfied response category; the remaining 68 households (36.2%) reported a neutral evaluation (Fig. 1).

Bars show the percentage of surveyed households in each evaluation category within each sampled jurisdiction; denominators are shown beside the bars. Satisfied and dissatisfied each combine the two corresponding response categories. Neutral responses are shown separately. Values are descriptive proportions from a convenience sample rather than county population estimates. County sample sizes match Table 1. Evaluation categories are defined in Section 2.3.1.

Low payment standards were reported by 1,588 households, representing 96.2% of the full sample (Table 2, panel A). Within this group, 614 households (38.7%) were satisfied with compensation overall, 523 (32.9%) were neutral and 451 (28.4%) were dissatisfied. Thus, 1,137 households (71.6%) reported low payment standards while giving a satisfied or neutral overall evaluation. A shared concern about payment standards therefore coexisted with different overall evaluations, rather than being confined to households expressing dissatisfaction.

Among households reporting low payment standards, overall dissatisfaction varied by maintenance arrangement (Table 2, panel B). Dissatisfaction was reported by 242 of 1,130 households under township or village collective maintenance (21.4%), compared with 137 of 287 under forest-farm maintenance (47.7%) and 72 of 170 under individual maintenance (42.4%). The single household reporting another maintenance arrangement was satisfied overall and provided insufficient observations for a group-level comparison. These descriptive subgroup comparisons show differences in overall evaluations among maintenance groups sharing the same reported concern.

Overall evaluations also differed by policy familiarity within the low-payment-standard subgroup (Table 2, panel C). Dissatisfaction was reported by eight of 120 very familiar respondents (6.7%), 351 of 1,290 basically familiar respondents (27.2%), 78 of 154 not-very-familiar respondents (50.6%), and 14 of 24 unfamiliar respondents (58.3%). The observed proportion increased across categories of decreasing

familiarity, although the unfamiliar category contained relatively few observations. These percentages describe the responses within each familiarity category in the low-payment-standard subgroup.

Table 2
Overall compensation evaluations alongside reported low payment standards. Panel A. All households by recorded compensation concern

Group	Households	Satisfied	Neutral	Dissatisfied
	n	n (%)	n (%)	n (%)
Low payment standard	1588	614 (38.7)	523 (32.9)	451 (28.4)
Other recorded reason	62	41 (66.1)	12 (19.4)	9 (14.5)
No recorded reason	0	—	—	—

Panel B. Low-standard subgroup by maintenance arrangement

Group	Households	Satisfied	Neutral	Dissatisfied
	n	n (%)	n (%)	n (%)
Collective	1130	471 (41.7)	417 (36.9)	242 (21.4)
Forest farm	287	89 (31.0)	61 (21.3)	137 (47.7)
Individual	170	53 (31.2)	45 (26.5)	72 (42.4)
Other	1	1 (100.0)	0 (0.0)	0 (0.0)

Panel C. Low-standard subgroup by policy familiarity

Group	Households	Satisfied	Neutral	Dissatisfied
	n	n (%)	n (%)	n (%)
Very familiar	120	103 (85.8)	9 (7.5)	8 (6.7)
Basically familiar	1290	486 (37.7)	453 (35.1)	351 (27.2)
Not very familiar	154	19 (12.3)	57 (37.0)	78 (50.6)
Unfamiliar	24	6 (25.0)	4 (16.7)	14 (58.3)

Notes: Entries are n (row %), except household counts. Panel A uses the single recorded Q7 concern: low payment standard = code 1; other recorded reason = codes 2–5. The question was available to all respondents, and no blank Q7 values occurred in the retained data. A dash denotes an undefined percentage for the empty category. Panels B and C describe the 1,588 households reporting low payment standards, by maintenance arrangement and self-reported policy familiarity, respectively.

Collective denotes township or village collective maintenance. These are descriptive subgroup comparisons.

3.3.2. Associations with Compensation Dissatisfaction in the Multivariable Logit Model

The main multivariable Logit model included 1,601 complete-case households from the full analytical sample, of whom 432 (27.0%) reported dissatisfaction. In this model, the odds of dissatisfaction were higher in Jingyuan (OR = 23.31, 95% CI: 7.66–70.98) and Pengyang (OR = 87.59, 95% CI: 15.47–495.95) than in Longde (Table 3; Fig. 2A). Both contrasts retained statistical support after Benjamini–Hochberg correction ($q < 0.001$; Table 3). The Yuanzhou–Longde contrast was less conclusive (OR = 2.07, 95% CI: 0.71–6.03). Confidence intervals were wide for Jingyuan and particularly Pengyang, with the latter represented by only one recorded locality (Table 1).

In the multivariable Logit model, individual maintenance was associated with higher odds of compensation dissatisfaction than township or village collective maintenance (OR = 3.24, 95% CI: 1.16–9.01; nominal P = 0.026; Table 3). However, the corresponding BH q value was 0.087, which did not meet the 0.05 threshold after multiplicity correction. The forest-farm comparison was imprecisely estimated (OR = 1.10, 95% CI: 0.15–8.28; nominal P = 0.923), providing no clear evidence of an association with dissatisfaction in this model.

In the multivariable Logit model, respondents who were not very familiar with the policy had higher odds of compensation dissatisfaction than those who were very familiar (OR = 11.82, 95% CI: 2.69–51.88; nominal P = 0.002; BH q = 0.010; Table 3). The estimates for basically familiar respondents (OR = 1.99, 95% CI: 0.78–5.10) and unfamiliar respondents (OR = 2.78, 95% CI: 0.35–22.27) were less conclusive, with both confidence intervals spanning one. The model estimates did not increase consistently across categories of decreasing familiarity. The model therefore supported the not-very-familiar versus very-familiar contrast, but did not establish a monotonic familiarity gradient.

The multivariable Logit model provided no clear evidence of a common linear association between household income category and the log odds of dissatisfaction (OR = 1.06 per category increase, 95% CI: 0.75–1.48; Table 3). The estimates for education (OR = 1.01 per category increase, 95% CI: 0.76–1.33) and age (OR = 1.27 per 10-year increase, 95% CI: 0.91–1.77) also had confidence intervals spanning one. The corresponding intervals for sex, receipt of forest-product income and household size likewise included one, and none of these six household characteristics met the BH $q < 0.05$ threshold. These results concern the associations estimated under the reported model specification and do not establish that household economic and demographic characteristics are unimportant to compensation evaluations.

Table 3
Associations with explicit compensation dissatisfaction in multivariable Logit models.

Comparison	All complete cases			Without Pengyang		
	OR (95% CI)	P	BH q	OR (95% CI)	P	BH q
Yuanzhou vs Longde	2.07 (0.71–6.03)	0.175	0.296	2.45 (0.94–6.42)	0.067	0.201
Jingyuan vs Longde	23.31 (7.66–70.98)	< 0.001	< 0.001	25.76 (8.85–74.97)	< 0.001	< 0.001
Pengyang vs Longde	87.59 (15.47–495.95)	< 0.001	< 0.001	—	—	—
Forest farm vs collective	1.10 (0.15–8.28)	0.923	0.991	0.90 (0.09–9.42)	0.924	0.991
Individual vs collective	3.24 (1.16–9.01)	0.026	0.087	2.41 (1.16–5.03)	0.021	0.087
Basically vs very familiar	1.99 (0.78–5.10)	0.146	0.296	1.92 (0.75–4.91)	0.165	0.296
Not very vs very familiar	11.82 (2.69–51.88)	0.002	0.010	10.27 (2.75–38.35)	0.001	0.008
Unfamiliar vs very familiar	2.78 (0.35–22.27)	0.323	0.435	3.17 (0.41–24.33)	0.257	0.385
Income category, +1	1.06 (0.75–1.48)	0.736	0.903	1.20 (0.83–1.74)	0.322	0.435
Education category, +1	1.01 (0.76–1.33)	0.970	0.991	1.00 (0.74–1.35)	0.991	0.991
Age, + 10 years	1.27 (0.91–1.77)	0.154	0.296	1.32 (0.94–1.85)	0.109	0.263
Male vs female	0.71 (0.46–1.08)	0.105	0.263	0.65 (0.45–0.94)	0.025	0.087

Notes: OR, odds ratio; CI, confidence interval; BH, Benjamini–Hochberg. Samples: all complete cases, n = 1,601 (432 dissatisfied households; 30 recorded localities); without Pengyang, n = 1,419 (316 dissatisfied households; 29 localities). Dissatisfied = 1; satisfied or neutral = 0. All listed variables entered simultaneously. References: Longde, collective maintenance, very familiar, female, and no forest-product income. Income and education are ascending category scores; age is per 10 years and household size per additional person. CIs and P values use HC1 locality-clustered covariance estimates with the cluster-count correction and t inference with G − 1 degrees of freedom. Nominal P values and 95% CIs are reported alongside BH q values, calculated jointly across the 27 non-intercept coefficients in the two Logit models. The correction does not cover all earlier exploratory analyses. A dash indicates that Pengyang was excluded. Complete coefficients and standard errors are provided in Supplementary Table S1.

Comparison	All complete cases			Without Pengyang		
	OR (95% CI)	P	BH q	OR (95% CI)	P	BH q
Forest income, yes vs no	0.07 (0.003–2.01)	0.117	0.263	1.83 (0.67–5.04)	0.231	0.366
Household size, +1	1.01 (0.90–1.14)	0.841	0.987	1.04 (0.92–1.18)	0.525	0.675

Notes: OR, odds ratio; CI, confidence interval; BH, Benjamini–Hochberg. Samples: all complete cases, n = 1,601 (432 dissatisfied households; 30 recorded localities); without Pengyang, n = 1,419 (316 dissatisfied households; 29 localities). Dissatisfied = 1; satisfied or neutral = 0. All listed variables entered simultaneously. References: Longde, collective maintenance, very familiar, female, and no forest-product income. Income and education are ascending category scores; age is per 10 years and household size per additional person. CIs and P values use HC1 locality-clustered covariance estimates with the cluster-count correction and t inference with G − 1 degrees of freedom. Nominal P values and 95% CIs are reported alongside BH q values, calculated jointly across the 27 non-intercept coefficients in the two Logit models. The correction does not cover all earlier exploratory analyses. A dash indicates that Pengyang was excluded. Complete coefficients and standard errors are provided in Supplementary Table S1.

3.4. Compensation Preferences and Sensitivity Analyses

3.4.1. Preferred Compensation Arrangements

Central government earmarked funds were the most frequently selected funding source (1,023 households; 62.0%), followed by local government earmarked funds (455; 27.6%). Central government funds ranked first in each sampled jurisdiction, with selection rates ranging from 47.5% in Longde to 79.7% in Yuanzhou. Donations from individuals and enterprises were selected by 37 households (2.2%), ecological taxes by 32 (1.9%), and higher water tariffs by 26 (1.6%); other sources were selected by 91 households (5.5%). No funding-source option was recorded for 609 households (36.9%). Multiple selections were permitted, and percentages were calculated using all 1,650 households, rather than the total number of selections (Supplementary Table S4, panel A).

Cash payments were selected by 1,419 households (86.0%) in the multiple-response question on compensation arrangements. Other selections included employment opportunities (470; 28.5%), payments calculated according to cultivated-land area (348; 21.1%), vocational training (329; 19.9%) and project investment (287; 17.4%). Preferential loan terms were selected by 165 households (10.0%), technical advisory services by 66 (4.0%), and other arrangements by 17 (1.0%). Respondents could select multiple arrangements, and all percentages used the full sample of 1,650 households. No compensation-arrangement option was recorded for 113 households (6.8%) (Supplementary Table S4, panel B).

Cash payments were the most frequently selected arrangement in all four sampled jurisdictions, with selection rates ranging from 77.1% in Pengyang to 91.4% in Jingyuan. Selections of other arrangements differed across locations. In Longde, 329 households (39.7%) selected employment opportunities and 227 (27.4%) selected project investment. In Yuanzhou, 121 households (40.9%) selected vocational

training and 125 (42.2%) selected payments calculated according to cultivated-land area. No employment, vocational-training or project-investment selections were recorded in Pengyang, where 43 of 188 households (22.9%) had no compensation-arrangement option recorded. These percentages use the full sample of each jurisdiction (Supplementary Table S4, panel B).

3.4.2. Sensitivity Analyses

Excluding Pengyang yielded a regression sample of 1,419 households, including 316 reporting dissatisfaction. In the model without Pengyang, the Jingyuan–Longde contrast remained positive (OR = 25.76, 95% CI: 8.85–74.97; BH $q < 0.001$), as did the not-very-familiar versus very-familiar contrast (OR = 10.27, 95% CI: 2.75–38.35; BH $q = 0.008$). For individual versus collective maintenance, the OR decreased from 3.24 in the full model to 2.41 (95% CI: 1.16–5.03; nominal $P = 0.021$; BH $q = 0.087$). The directions of these three contrasts were retained, but the individual-maintenance comparison did not meet the BH $q < 0.05$ threshold in either sample (Table 3; Fig. 2A).

Estimates for household characteristics also varied between the two samples (Table 3). The OR per one-category increase in household income changed from 1.06 (95% CI: 0.75–1.48) in the full model to 1.20 (95% CI: 0.83–1.74) without Pengyang; neither confidence interval excluded 1. The forest-product-income estimate changed direction, from an OR of 0.07 (95% CI: 0.003–2.01) to 1.83 (95% CI: 0.67–5.04). Both intervals included 1, and the full-model interval was particularly wide, providing no consistent evidence of the direction of this association. For male versus female respondents, the OR changed from 0.71 (95% CI: 0.46–1.08) to 0.65 (95% CI: 0.45–0.94). Although the comparison without Pengyang had a nominal P value of 0.025, its BH q value was 0.087 and therefore did not meet the $q < 0.05$ threshold.

In linear probability models with recorded-locality fixed effects, individual maintenance was associated with a 19.9-percentage-point higher probability of reported dissatisfaction than collective maintenance in the full sample (95% CI: 7.9–31.8 percentage points; BH $q = 0.014$). Without Pengyang, this difference was 14.8 percentage points (95% CI: 3.7–25.9; nominal $P = 0.011$; BH $q = 0.060$). Thus, both estimates were positive, but only the full-sample comparison met the BH $q < 0.05$ threshold. The forest-farm versus collective-maintenance estimates were 6.7 percentage points (95% CI: −19.9–33.2) in the full sample and 2.5 percentage points (95% CI: −31.0–36.0) without Pengyang, with neither interval excluding zero (Supplementary Table S1; Fig. 2B).

The not-very-familiar versus very-familiar contrast also persisted in the recorded-locality fixed-effects models. The estimated difference in the probability of reported dissatisfaction was 30.3 percentage points in the full sample (95% CI: 16.3–44.2; BH $q = 0.0013$) and 29.7 percentage points without Pengyang (95% CI: 17.4–42.0; BH $q < 0.001$). For basically familiar versus very familiar, the corresponding differences were 4.1 percentage points (95% CI: −2.9–11.1) and 4.0 percentage points (95% CI: −2.6–10.6). For unfamiliar versus very familiar, the estimates were larger but less precise: 35.4 percentage points (95% CI: −0.7–71.4) and 33.8 percentage points (95% CI: −3.1–70.6), respectively. Statistical support was therefore concentrated in the not-very-familiar contrast; the other two category comparisons did not meet the BH $q < 0.05$ threshold in either sample (Supplementary Table S1; Fig. 2B).

Panel A shows ORs from multivariable Logit models on a logarithmic scale; the dashed line marks OR = 1. Panel B shows coefficients from linear probability models with recorded-locality fixed effects in percentage points; the dashed line marks zero. The two panels use different effect scales and should not be compared numerically. All models include maintenance arrangements, policy familiarity, income and education category scores, age, sex, forest-product income and household size. Panel A additionally includes county indicators; Panel B replaces these with recorded-locality fixed effects. Samples comprise 1,601 complete cases (30 localities) or 1,419 complete cases (29 localities) without Pengyang. Error bars show nominal 95% CIs using HC1 locality-clustered covariance estimates, the cluster-count correction and t critical values with G − 1 degrees of freedom. Intervals alone do not determine whether a comparison meets the BH q < 0.05 threshold. BH q values were calculated jointly across the 27 non-intercept coefficients in the two Logit models and, separately, across the 22 maintenance, familiarity and household coefficients in the two linear probability models. These families do not cover all earlier exploratory analyses. Selected contrasts are displayed; complete coefficients, standard errors, P values and q values are reported in Supplementary Table S1, with Logit results also summarized in Table 3.

4. Discussion

4.1. Payment concerns and overall compensation evaluations

A widespread concern about payment levels coexisted with markedly different overall evaluations of compensation. Although 96.2% of the surveyed households reported low payment standards, 71.6% of this subgroup were satisfied or neutral overall, and 28.4% were dissatisfied (Table 2). Payment-specific concerns and overall evaluations therefore provide complementary information: the former identifies a commonly reported issue, whereas the latter reveals variation among households sharing that concern. This finding complements the Jinghe River Basin study by Wang et al. (2026), which examined residents' perceptions and used contingent valuation to estimate willingness to accept or pay for multifunctional forest ecological compensation. Similarly, Lindhjem and Mitani (2012) used contingent valuation to elicit Norwegian private forest owners' willingness to accept compensation for voluntary conservation. Whereas monetary preference elicitation asks what respondents would be willing to accept or pay under a specified scenario, the paired responses in our survey show how a concern about payment levels can coexist with different evaluations of existing compensation. Establishing whether these concerns correspond to an objective shortfall requires a further comparison of actual payments with relevant costs. For example, Zheng et al. (2013) integrated household and hydrological data to evaluate a watershed payment programme's costs, benefits and livelihood implications. Our findings thus identify households and questions for follow-up, rather than a compensation rate that can be inferred from the prevalence of low-standard responses alone.

Different evaluation criteria offer one plausible explanation for these response combinations. A household might compare the payment amount with its perceived costs or expectations, while

evaluating compensation overall in terms of the continuity of support or recognition of its conservation efforts. These possibilities are consistent with Chapman et al. (2020), who found that many participants in Costa Rica's PES programme interpreted payments as support for land stewardship rather than solely as transactions for specified services. However, such interpretations were not directly measured in our survey. The response process offers a competing explanation. Schwarz (1999) explained how question wording, response formats and context can shape self-reports, while Clot et al. (2017) examined how describing a proposed PES incentive as a payment or compensation affected programme evaluations. In our survey, the reason question was available to all households, including those satisfied or neutral overall. Although respondents could leave it unanswered, every retained record contained a reason code. This pattern warrants checking whether respondents understood the item as asking them to identify an aspect requiring improvement, regardless of their overall evaluation. Village cadres' presence also raises questions about the perceived privacy of policy evaluations; research on sensitive questions identifies reporting conditions as a potential source of response error(Tourangeau and Yan 2007). This does not establish that cadre presence altered answers in our survey. The observed combinations therefore support distinguishing payment concerns from overall evaluations, but cannot establish whether their coexistence arose from different evaluation criteria, item interpretation, reporting conditions or a combination of these processes.

Targeted follow-up could clarify which comparison standards and question interpretations underlie these response combinations. Questionnaire-evaluation research provides methods for examining how survey items function in practice(Presser et al. 2004), while cognitive interviewing collects respondents' explanations of how they understood questions and arrived at their answers(Beatty and Willis 2007). Applied here, households reporting low payment standards could be purposively selected from each overall-evaluation category across surveyed localities. In private interviews without village cadres present, respondents would first answer the original items and then explain, in their own words, what they understood by low payment standards and overall satisfaction. Subsequent probes could explore the payments, costs or expectations used as comparison standards, the aspects considered in the overall evaluation, and how respondents decided whether the reason item required an answer. Recurring interpretation problems could inform revised wording and a clearly stated option to report no concern, followed by pretesting before a new survey. This process would help distinguish substantive differences in evaluation criteria from difficulties with question interpretation. It would not estimate the population prevalence of either explanation, and changed answers at follow-up alone would not isolate an effect of cadre presence from changes in timing or survey mode.

4.2. Policy familiarity and household livelihood conditions

Policy familiarity provides a useful diagnostic signal because the not-very-familiar versus very-familiar contrast persisted across the reported model specifications. The main model used all 1,601 complete-case households, including households with other recorded concerns. In the main Logit model, respondents reporting that they were not very familiar had higher odds of dissatisfaction (OR = 11.82, 95% CI: 2.69–51.88; BH q = 0.010; Table 3), and this contrast remained supported after excluding

Pengyang and in linear probability models with recorded-locality fixed effects (Fig. 2). However, the category estimates did not establish a monotonic familiarity gradient. Moreover, the item measured respondents' perceived familiarity, rather than demonstrated knowledge of payment rules or exposure to a specified communication intervention. Related PES research helps explain why this distinction matters: household preferences can depend on contractual conditions(Li et al. 2017), while field evidence from the Sloping Land Conversion Program highlights the importance of implementation arrangements and local participation(He and Lang 2015). Familiarity in our sample could therefore reflect differences in access to officials or experience with programme implementation, alongside differences in policy understanding. Dissatisfaction might also influence how respondents assess their own familiarity. These are competing explanations, rather than mechanisms identified by the cross-sectional models. Communication itself is not a uniform intervention: in a residential water-use experiment, technical information, prosocial appeals and social-comparison messages produced different behavioural responses(Ferraro and Price 2013). Although that outcome differs from compensation satisfaction, the study illustrates why the content and effects of a communication strategy require direct evaluation. The practical implication is to investigate what less-familiar households understand and have experienced before deciding whether additional information is the appropriate response to their dissatisfaction.

Understanding compensation rules does not necessarily imply accepting them. A household may accurately understand its entitlement and receive the stipulated payment, yet still consider the payment standard or allocation procedure unacceptable. Conversely, incomplete understanding may coexist with delayed or incomplete payments. PES research provides a basis for distinguishing these situations: Pascual et al. (2014) argue for integrating social equity into programme design and evaluation, while Corbera et al. (2007) show how participation, information distribution and local social relations can shape the legitimacy of forest carbon projects. These perspectives suggest that dissatisfaction should prompt separate examination of policy understanding, payment delivery and acceptance of the rules. In our sample, 84.8% of households reported government publicity as an information source (Supplementary Table S3), but this indicates a reported communication channel, not verified comprehension or fulfilment of payment entitlements. Follow-up should therefore compare respondents' accounts of payment rules and obligations with the provisions applicable to their enrolled land, and compare recorded payment amounts and dates with those entitlements. Separate questions should examine how households judge payment levels, eligibility criteria and opportunities to raise concerns. Clearer explanations may address misunderstandings; discrepancies between entitlements and receipts require implementation checks; and objections to correctly understood rules require substantive consideration rather than an assumption that further explanation will secure agreement. These situations may overlap within the same household. The present survey cannot determine their respective contributions to dissatisfaction and does not establish that payment failures or inequitable allocation occurred among the surveyed households.

Total household income does not fully describe the livelihood conditions under which compensation is evaluated. The income-category estimate in our main model provided no clear evidence of a common trend in dissatisfaction across ascending categories (OR = 1.06; 95% CI: 0.75–1.48; Table 3). This

specification does not test whether agricultural dependence, employment stability or competing demands on household labour distinguish evaluations among households in the same income category. The income-source question recorded activities such as farming and temporary employment, but not their respective shares of household earnings (Supplementary Table S2). Earlier studies demonstrate why these distinctions merit investigation. Li et al. (2011) found no increase in labour transfer towards nonfarm activities at their survey site, despite reported income gains from programme participation. Uchida et al. (2009) found that shifts towards off-farm employment varied with initial human and physical capital, with evidence consistent with relaxation of liquidity constraints. By contrast, Kelly and Huo attributed increased nonfarm employment in their study to substitution of farm labour into nonfarm work rather than alleviation of constraints. These different explanations caution against assuming a single livelihood response to conservation payments. For Guyuan, a testable hypothesis is that households in similar income categories may evaluate compensation differently because their reliance on land-based earnings, maintenance obligations and opportunities to earn income elsewhere differ. Neither the income coefficient nor the presence of forest-product income directly tests this hypothesis; the latter records receipt rather than dependence, and its estimate was unstable across samples (Table 3). Testing this explanation would require source-specific income amounts, the share derived from compensation, maintenance expenditure and labour requirements, together with information on access to alternative employment. The current findings leave this hypothesis open; they neither establish it nor show that economic constraints are unimportant.

4.3. Maintenance arrangements and geographical variation

Maintenance arrangements provide a practical starting point for investigating who undertakes conservation work and what support accompanies that responsibility. Individual maintenance was positively associated with dissatisfaction in the multivariable Logit model (OR = 3.24; nominal P = 0.026), although the comparison did not meet the BH $q < 0.05$ threshold ($q = 0.087$; Table 3). The direction remained positive in the recorded-locality fixed-effects models, but the BH q value changed from 0.014 in the full sample to 0.060 after excluding Pengyang (Fig. 2B). This pattern warrants investigation of implementation arrangements, while showing that the strength of evidence depends on specification and sample. The literature identifies concrete dimensions for such investigation. Falconer (2000) found that programme-related transaction costs can constrain farmers' participation in European agri-environmental schemes. In China's cropland-to-forest programme, Bennett et al. (2014) reported relationships between household labour availability, opportunity costs, local implementation arrangements and reported tree survival. These studies concern participation and tree survival, respectively; they do not demonstrate that individual maintenance causes dissatisfaction or necessarily imposes higher costs in Guyuan. Our maintenance variable records the reported arrangement, not labour days, expenditure, the allocation of specific duties or access to technical support. Collective or forest-farm maintenance likewise does not establish that households bear no responsibilities or costs. Follow-up should compare actual duties and inputs across arrangements, record how responsibilities were assigned, and account for differences in parcel conditions and available support. Such evidence would help assess whether dissatisfaction relates to uncompensated work, unclear responsibilities or other

implementation concerns. The present association alone provides no basis for concluding that replacing individual maintenance with collective maintenance would improve household evaluations.

County contrasts identify where household evaluations differed, but do not explain why they differed. The higher odds of dissatisfaction in Jingyuan and Pengyang relative to Longde (OR = 23.31 and 87.59, respectively; Table 3) justify closer examination of the sampled locations. Their magnitude should nevertheless be read alongside the wide confidence intervals and uneven geographical coverage: Jingyuan and Pengyang were represented by three and one recorded localities, respectively, compared with ten in Longde and sixteen in Yuanzhou (Table 1). In particular, the Pengyang contrast cannot separate a general county pattern from circumstances specific to its single sampled locality. Earlier SLCP research identifies possible implementation issues to investigate. Bennett(2008) documented subsidy shortfalls, insufficient technical support and limited consultation in earlier programme implementation, while He (2021) examined how formal programme arrangements interacted with pre-existing local institutions. These findings support attention to local implementation processes, without establishing that the same problems occurred in our survey locations. Moreover, Zinda and Zhang (2019) showed that community-level vegetation outcomes depended on combinations of governance, livelihoods and environmental conditions; responsive governance did not have a consistent relationship with forest gain. Although their outcome differs from household evaluations, this cautions against treating geographical variation as a simple ranking of administrative performance. For Guyuan, comparisons across and within counties should examine applicable payment rules, amounts and timing of receipts, policy communication, maintenance responsibilities, livelihood conditions and survey administration. County indicators may capture several of these differences, alongside other unmeasured characteristics. The resulting evidence could identify problems requiring local responses, but the present ORs measure relative odds of dissatisfaction, not monetary losses, ecosystem-service values or appropriate payment rates. They therefore provide neither a county governance score nor a basis for setting differentiated compensation amounts.

The sensitivity analyses strengthen the case for following up the Jingyuan–Longde and not-very-familiar versus very-familiar contrasts within the surveyed locations. After excluding Pengyang, both contrasts remained positive and met the BH $q < 0.05$ threshold, indicating that their statistical support did not depend solely on including that county (Table 3). The familiarity contrast also persisted in models with recorded-locality fixed effects (Fig. 2B). These models use within-locality variation while allowing each recorded locality its own intercept, reducing the scope for shared differences between localities to account for that association. They do not remove unmeasured differences between households within a locality. Because these models also use a different link and effect scale from the Logit models, consistency should be assessed through direction and uncertainty rather than numerical equivalence. Evidence for individual maintenance was less consistent across specifications and samples, while the forest-product-income estimate changed direction, preventing a general claim that all associations were stable. Neither analysis tested performance in unobserved locations: excluding Pengyang involved refitting models to the remaining households, not evaluating predictions for the excluded households. Model-transferability research makes this distinction explicit. Wenger and Olden (2012) distinguish

conventional model assessment from testing transferability to other settings, while Roberts et al. (2017) explain why validation should account for spatial and hierarchical dependence. Evaluating geographical prediction would require independent locations or an appropriate geographically structured holdout design, using a model capable of predicting previously unseen locations. Such evaluation would still not make a convenience sample representative of wider populations. The present contribution is therefore greater confidence in selected follow-up priorities within the surveyed setting, rather than a validated prediction rule for other counties.

4.4. Implications for compensation diagnosis and programme evaluation

The policy implication is to identify what requires correction before choosing how compensation should be redesigned. The coexistence of a widely reported concern about low payment standards and differing overall evaluations motivates a proposed three-stage diagnostic procedure. PES design research emphasizes the importance of programme design for effectiveness and distributional outcomes(Engel et al. 2008). In the first stage, applicable payment entitlements should be checked against administrative records and household accounts of amounts and timing of receipts, alongside short questions assessing understanding of the relevant rules. Existing payment obligations should be fulfilled independently of any pilot evaluation. The second stage would examine concerns that remain after these checks, documenting maintenance expenditure, unpaid labour, foregone land returns and constraints on alternative livelihoods, together with households' assessments of the rules and their application. These inquiries would test whether particular concerns correspond to identifiable delivery problems, information gaps, costs or livelihood constraints, rather than assigning causes from satisfaction responses alone. In the third stage, measures matched to verified problems could be piloted and revised using monitored outcomes and participant feedback. Adaptive changes to environmental and social targeting in Mexico's hydrological-services programme illustrate learning from implementation, although they do not validate the procedure proposed here(Sims et al. 2014). Evaluation should separately track payment delivery, policy understanding, livelihood outcomes and land maintenance, while retaining satisfaction as a distinct subjective outcome. Where ecosystem-service benefits are claimed, relevant biophysical measurements would also be required. This distinction is consistent with the emphasis on balancing forest ecosystem-service trade-offs and differing stakeholder priorities in Asbjornsen et al. (2022). The procedure is a recommendation for trial and evaluation, not a compensation-pricing rule or an intervention whose effectiveness has been established by this survey.

The preference results identify options for follow-up, but do not support replacing cash compensation with livelihood programmes. Cash payments were selected by 86.0% of surveyed households, while employment opportunities, vocational training and project investment were selected by 28.5%, 19.9% and 17.4%, respectively (Supplementary Table S4). Because respondents could select several options, these percentages do not reveal how much cash they would exchange for another form of support, whether they would participate under specified conditions, or whether any option would reduce dissatisfaction. The importance of testing such trade-offs is illustrated by Costedoat et al. (2016): their

choice experiment in Chiapas, Mexico, found a strong preference for individual cash payments and reluctance to allocate payments to collective investments. That finding does not establish preferences in Guyuan, but cautions against assuming that development projects are acceptable substitutes for cash. In our sample, the relatively frequent selection of employment opportunities in Longde and vocational training in Yuanzhou identifies locally relevant topics for consultation, rather than ready-made county intervention packages. Moreover, payments calculated according to cultivated-land area describe a payment basis and should not be interpreted as a distinct non-cash benefit. These distinctions accord with Jack et al. (2008), who explain how environmental, socioeconomic and political contexts interact with PES design to shape outcomes. Follow-up choice tasks could therefore specify payment amounts, support quality, time commitments and eligibility conditions, with an option to retain the existing arrangement. Subsequent pilots could assess actual uptake, access and livelihood outcomes. Existing cash entitlements would remain separate from these tests, and neither uptake nor improved satisfaction alone would demonstrate ecological benefits.

Fiscal appraisal should compare the additional costs and outcomes of responses matched to verified problems. The present survey does not provide the expenditure and counterfactual outcome data needed to establish that current compensation spending is inefficient. Nor does interest in employment or project investment demonstrate that these alternatives would deliver better value for money. Ferraro and Simpson (2002) show that direct conservation payments can be more cost-effective than interventions that support commercial activities to generate conservation indirectly, cautioning against assuming that a broader development package is necessarily more efficient. Spatial targeting research similarly considers expected additional environmental benefits and participation costs(Wünscher et al. 2008); dissatisfaction rates and county odds ratios do not supply these quantities. For Guyuan, fiscal evaluation of diagnostic pilots should record payments and the additional costs of diagnosis, administration, support and monitoring, while a separate social assessment should account for households' labour and other opportunity costs without double counting. Baseline and follow-up measurements should be combined with a credible comparison design, such as randomized introduction of additional voluntary support where feasible, without withholding existing payment entitlements. Outcomes should correspond to the problem being addressed: for example, verified payment timeliness for delivery interventions, tested understanding for information interventions, and net livelihood gains for employment support. Claims about ecological cost-effectiveness would additionally require evidence of ecological gains relative to what would have occurred without the intervention. Indicator selection also requires care: Zabel and Roe (2009) identify risks outside participants' control and distortion in environmental-service measurement as issues in performance-payment design. Decisions to expand a pilot should consider its incremental costs, demonstrated outcomes and the distribution of benefits and burdens across household groups.

5. Conclusions and Limitations

5.1. Main Conclusions

(1) Overall compensation evaluations differed markedly across the surveyed households and locations. Among the 1,650 Grain-for-Green households in Guyuan, 39.7% were satisfied, 32.4% were neutral and 27.9% were dissatisfied. Dissatisfaction ranged from 8.3% in the Longde sample to 65.9% in Jingyuan, showing that the overall sample average concealed substantial geographical differences.

(2) Considering payment standards low did not necessarily mean being dissatisfied with compensation overall. Among the 1,588 households reporting low standards, 38.7% were satisfied, 32.9% were neutral and 28.4% were dissatisfied. A concern about the payment amount therefore cannot substitute for an overall evaluation or, by itself, demonstrate an objective compensation gap.

(3) Location and a specific policy-familiarity contrast showed more consistent associations with dissatisfaction than the other examined characteristics. In the multivariable models, Jingyuan households had higher odds of dissatisfaction than Longde households, and respondents who were not very familiar with policy had higher odds than those who were very familiar. These contrasts persisted after excluding Pengyang, and the familiarity contrast also persisted in models with recorded-locality fixed effects. Evidence for individual maintenance was less consistent, while total income category showed no clear common linear association under the reported specification. These findings identify priorities for investigation rather than established causes of dissatisfaction.

(4) Cash was the most frequently selected compensation arrangement (86.0%), while employment opportunities (28.5%) and vocational training (19.9%) were also selected. These multiple-response preferences identify options for consultation. Policy responses should begin with verification of payment delivery and policy understanding, followed by assessment of household costs and livelihood constraints, and then pilots matched to verified problems. The proposed procedure requires evaluation through separate monitoring of payment delivery, understanding, livelihoods and land maintenance.

5.2. Limitations and Future Research Directions

This study has several limitations. First, convenience sampling, uneven geographical coverage and the presence of village cadres during questionnaire completion may have affected sample composition and responses, limiting generalization beyond the surveyed households. Second, the cross-sectional data identify associations rather than causal relationships, and self-reported familiarity and maintenance categories do not directly measure policy knowledge or household costs. Third, regression excluded 49 households, among whom dissatisfaction was more common than among retained households, creating a potential source of selection bias.

Future research should include more independently selected survey locations and provide private conditions for questionnaire completion. Follow-up interviews and repeated surveys could clarify question interpretation and changes in household evaluations. Linking responses with payment records, livelihood and maintenance costs, and ecological observations would help test competing explanations and evaluate measures matched to verified problems.

Declarations

CRediT authorship contribution statement: Miao Wu: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Writing – original draft. Zhijia Yu: Funding acquisition, Resources, Validation. Jie Ma: Investigation, Data curation, Visualization. Zhengan Wang: Conceptualization, Supervision, Project administration, Writing – review & editing.

Declaration of competing interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Conflicts of interest:

The authors declare no conflicts of interest.

Funding:

The research was supported by the 2024 Youth Science and Technology Talent Support Project of Ningxia Hui Autonomous Region (Grant No. Ningkexie [2025] 2), the Engineering Technology Research Center for Protection and Breeding of Unique Wild Forest Resources in Liupan Mountains, Ningxia Hui Autonomous Region (Grant No. 2023GCJS012), and the sub-project of the Joint Fund of the National Natural Science Foundation of China (Grant No. U20A2085).

Data availability Statement:

The original survey data presented in this study are not publicly available due to privacy and ethical restrictions. Data supporting the reported results can be obtained from the corresponding author upon reasonable request.

References

1. Asbjornsen H, Wang Y, Ellison D et al (2022) Multi-Targeted payments for the balanced management of hydrological and other forest ecosystem services. For Ecol Manag 522:120482. https://doi.org/10.1016/j.foreco.2022.120482
2. Beatty PC, Willis GB (2007) Research Synthesis: The Practice of Cognitive Interviewing. Pub Opin Q 71:287–311. https://doi.org/10.1093/poq/nfm006
3. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Royal Stat Soc Ser B: Stat Methodol 57:289–300.

https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

4. Bennett M, Xie C, Hogarth N et al (2014) China's Conversion of Cropland to Forest Program for Household Delivery of Ecosystem Services: How Important is a Local Implementation Regime to Survival. Rate Outcomes? Forests 5:2345–2376. https://doi.org/10.3390/f5092345
5. Bennett MT (2008) China's sloping land conversion program: Institutional innovation or business as usual? Ecol Econ 65:699–711. https://doi.org/10.1016/j.ecolecon.2007.09.017
6. Chapman M, Satterfield T, Wittman H, Chan KMA (2020) A payment by any other name: Is Costa Rica's PES a payment for services or a support for stewards? World Dev 129:104900. https://doi.org/10.1016/j.worlddev.2020.104900
7. Clot S, Grolleau G, Méral P (2017) Payment Vs. Compensation For Ecosystem Services: Do Words Have A Voice. Des Environ Conserv Programs? Ecol Econ 135:299–303. https://doi.org/10.1016/j.ecolecon.2016.12.028
8. Corbera E, Brown K, Adger WN (2007) The Equity and Legitimacy of Markets for Ecosystem Services. Dev Change 38:587–613. https://doi.org/10.1111/j.1467-7660.2007.00425.x
9. Costedoat S, Koetse M, Corbera E, Ezzine-de-Blas D (2016) Cash only? Unveiling preferences for a PES contract through a choice experiment in Chiapas, Mexico. Land Use Policy 58:302–317. https://doi.org/10.1016/j.landusepol.2016.07.023
10. Engel S, Pagiola S, Wunder S (2008) Designing payments for environmental services in theory and practice: An overview of the issues. Ecol Econ 65:663–674. https://doi.org/10.1016/j.ecolecon.2008.03.011
11. Falconer K (2000) Farm-level constraints on agri-environmental scheme participation: a transactional perspective. J Rural Stud 16:379–394. https://doi.org/10.1016/S0743-0167(99)00066-2
12. Ferraro PJ, Price MK (2013) Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment. Rev Econ Stat 95:64–73. https://doi.org/10.1162/REST_a_00344
13. Ferraro PJ, Simpson RD (2002) The Cost-Effectiveness of Conservation Payments. Land Econ 78:339–353. https://doi.org/10.2307/3146894
14. General Office of Guyuan Municipal People's Government (2024) Master plan for national forest city development in Guyuan, Ningxia (2022–2031). General Office of Guyuan Municipal People's Government, Guyuan, China
15. General Office of Guyuan Municipal People's Government (2022) Natural resource protection and utilization plan for Guyuan during the 14th Five-Year Plan period. General Office of Guyuan Municipal People's Government, Guyuan, China
16. General Office of Guyuan Municipal People's Government (2023) Medium- and long-term special plan for urban and rural household waste classification and harmless treatment in Guyuan (2023–2035). General Office of Guyuan Municipal People's Government, Guyuan, China
17. Guyuan Municipal Bureau of Statistics (2025) Analysis of economic performance in Guyuan in 2024. Guyuan Municipal Bureau of Statistics, Guyuan, China

18. He J (2021) Institutional bricolage in payment for ecosystem services: Insights from the Sloping Land Conversion Programme in upland communities, Southwest China. Asia Pac Viewp 62:164–178. https://doi.org/10.1111/apv.12298
19. He J, Lang R (2015) Limits of State-Led Programs of Payment for Ecosystem Services: Field Evidence from the Sloping Land Conversion Program in Southwest China. Hum Ecol 43:749–758. https://doi.org/10.1007/s10745-015-9782-9
20. Jack BK, Kousky C, Sims KRE (2008) Designing payments for ecosystem services: Lessons from previous experience with incentive-based mechanisms. Proc Natl Acad Sci USA 105:9465–9470. https://doi.org/10.1073/pnas.0705503104
21. Kelly P, Huo X (2013) Land Retirement and Nonfarm Labor Market Participation: An Analysis of China's Sloping Land Conversion Program. World Dev 48:156–169. https://doi.org/10.1016/j.worlddev.2013.04.002
22. Li H, Bennett MT, Jiang X et al (2017) Rural Household Preferences for Active Participation in Payment for Ecosystem Service Programs: A Case in the Miyun Reservoir Catchment, China. PLoS ONE 12:e0169483. https://doi.org/10.1371/journal.pone.0169483
23. Li J, Feldman MW, Li S, Daily GC (2011) Rural household income and inequality under the Sloping Land Conversion Program in western China. Proc Natl Acad Sci USA 108:7721–7726. https://doi.org/10.1073/pnas.1101018108
24. Lindhjem H, Mitani Y (2012) Forest owners' willingness to accept compensation for voluntary conservation: A contingent valuation approach. JFE 18:290–302. https://doi.org/10.1016/j.jfe.2012.06.004
25. Pascual U, Phelps J, Garmendia E et al (2014) Social Equity Matters in Payments for Ecosystem Services. Bioscience 64:1027–1036. https://doi.org/10.1093/biosci/biu146
26. Presser S, Couper MP, Lessler JT et al (2004) Methods for Testing and Evaluating Survey Questions. Pub Opin Q 68:109–130. https://doi.org/10.1093/poq/nfh008
27. Roberts DR, Bahn V, Ciuti S et al (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40:913–929. https://doi.org/10.1111/ecog.02881
28. Schwarz N (1999) Self-reports: How the questions shape the answers. Am Psychol 54:93–105. https://doi.org/10.1037/0003-066X.54.2.93
29. Sims KRE, Alix-Garcia JM, Shapiro-Garza E et al (2014) Improving Environmental and Social Targeting through Adaptive Management in Mexico's Payments for Hydrological Services Program. Conserv Biol 28:1151–1159. https://doi.org/10.1111/cobi.12318
30. Tourangeau R, Yan T (2007) Sensitive questions in surveys. Psychol Bull 133:859–883. https://doi.org/10.1037/0033-2909.133.5.859
31. Uchida E, Rozelle S, Xu J (2009) Conservation Payments, Liquidity Constraints, and Off-Farm Labor: Impact of the Grain-for-Green Program on Rural Households in China. Am J Agri Econ 91:70–86. https://doi.org/10.1111/j.1467-8276.2008.01184.x

32. Wang X, Shu L, Wang Y (2026) Residents' Perceptions and Willingness to Pay for Multifunctional Ecological Compensation in Watershed Forests: Evidence from the Jinghe River Basin, the Loess Plateau. Forests 17:189. https://doi.org/10.3390/f17020189
33. Wenger SJ, Olden JD (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol Evol 3:260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x
34. Wünscher T, Engel S, Wunder S (2008) Spatial targeting of payments for environmental services: A tool for boosting conservation benefits. Ecol Econ 65:822–833. https://doi.org/10.1016/j.ecolecon.2007.11.014
35. Zabel A, Roe B (2009) Optimal design of pro-conservation incentives. Ecol Econ 69:126–134. https://doi.org/10.1016/j.ecolecon.2009.08.001
36. Zeileis A, Köll S, Graham N (2020) Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in *R*. J Stat Soft 95. https://doi.org/10.18637/jss.v095.i01
37. Zheng H, Robinson BE, Liang Y-C et al (2013) Benefits, costs, and livelihood implications of a regional payment for ecosystem service program. Proc Natl Acad Sci USA 110:16681–16686. https://doi.org/10.1073/pnas.1312324110
38. Zinda JA, Zhang Z (2019) Explaining heterogeneous afforestation outcomes: How community officials and households mediate tree cover change in China. World Dev 122:385–398. https://doi.org/10.1016/j.worlddev.2019.05.020

Figures

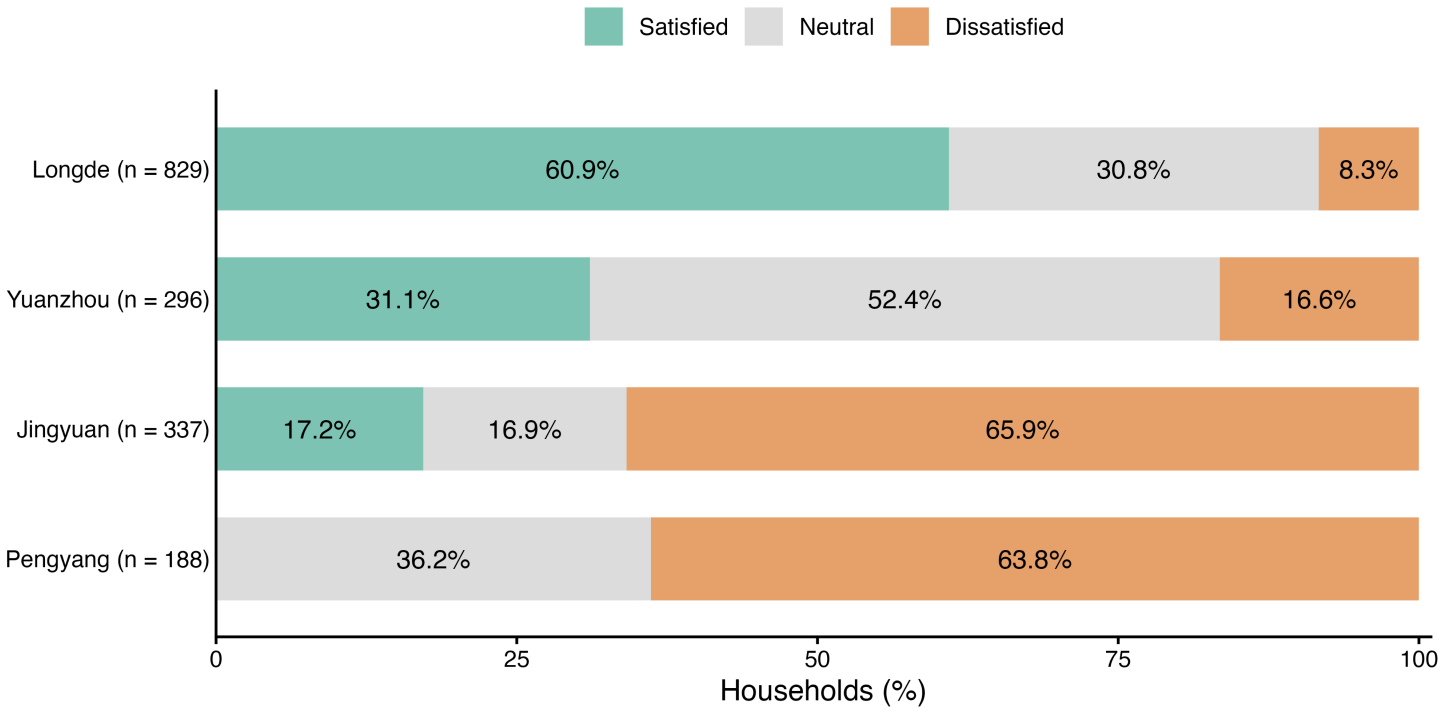


Figure 1

Distribution of overall compensation evaluations by county-level jurisdiction.

Bars show the percentage of surveyed households in each evaluation category within each sampled jurisdiction; denominators are shown beside the bars. Satisfied and dissatisfied each combine the two corresponding response categories. Neutral responses are shown separately. Values are descriptive proportions from a convenience sample rather than county population estimates. County sample sizes match Table 1. Evaluation categories are defined in Section 2.3.1.

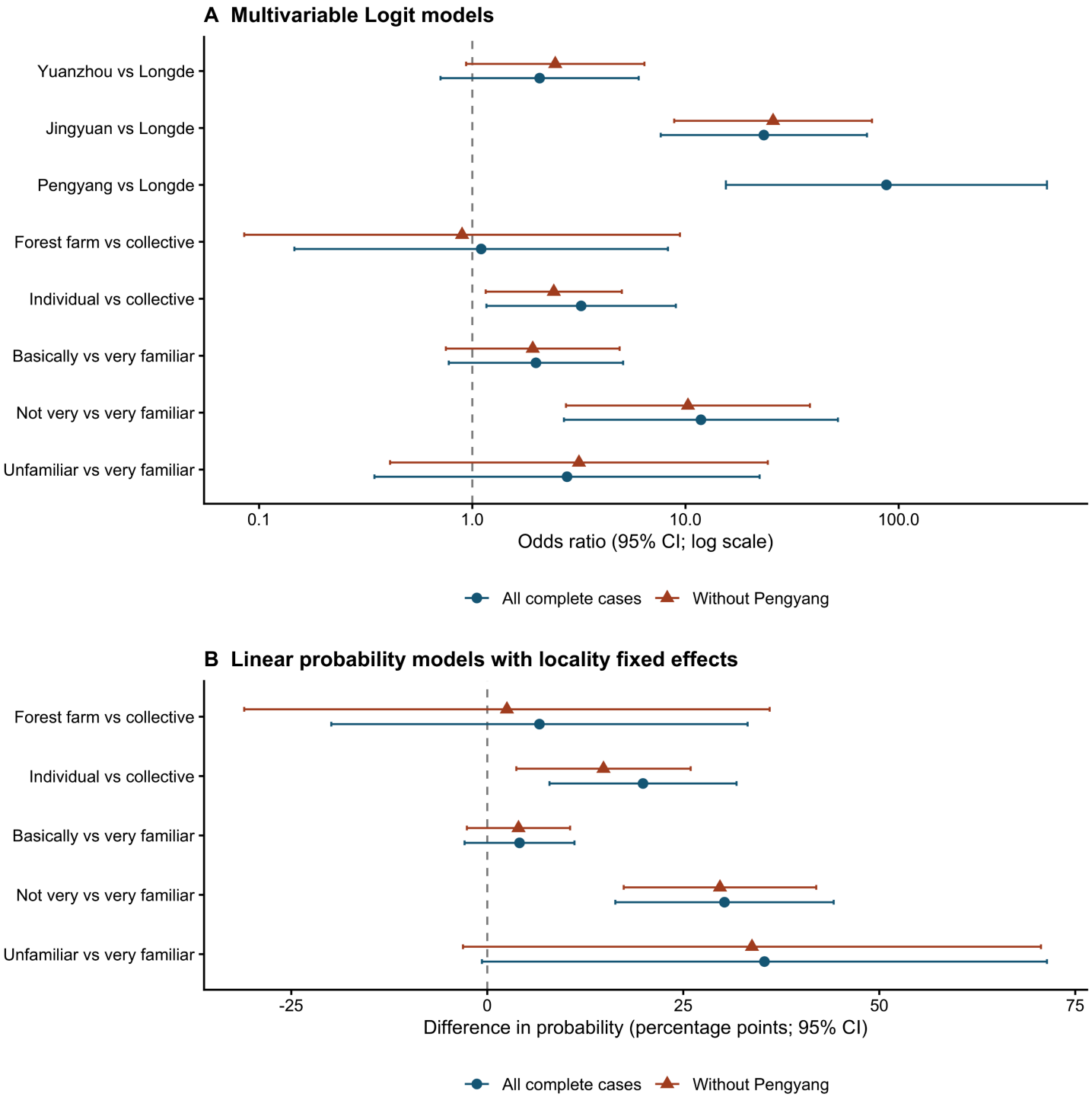


Figure 2

Selected model associations and sensitivity to exclusion of Pengyang.

Panel A shows ORs from multivariable Logit models on a logarithmic scale; the dashed line marks OR = 1. Panel B shows coefficients from linear probability models with recorded-locality fixed effects in percentage points; the dashed line marks zero. The two panels use different effect scales and should not be compared numerically. All models include maintenance arrangements, policy familiarity, income and education category scores, age, sex, forest-product income and household size. Panel A additionally includes county indicators; Panel B replaces these with recorded-locality fixed effects. Samples comprise 1,601 complete cases (30 localities) or 1,419 complete cases (29 localities) without Pengyang. Error bars show nominal 95% CIs using HC1 locality-clustered covariance estimates, the cluster-count correction and t critical values with G − 1 degrees of freedom. Intervals alone do not determine whether a comparison meets the BH q < 0.05 threshold. BH q values were calculated jointly across the 27 non-intercept coefficients in the two Logit models and, separately, across the 22 maintenance, familiarity and household coefficients in the two linear probability models. These families do not cover all earlier exploratory analyses. Selected contrasts are displayed; complete coefficients, standard errors, P values and q values are reported in Supplementary Table S1, with Logit results also summarized in Table 3.