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.

A “One Species, Two Standards” Bidirectional Regulation Protocol Based on Ecological Niche Property Encoding

Guanfeng Yang

jjaarr@163.com

Foshan Yesheng Health Consulting Co., Ltd. https://orcid.org/0009-0009-4933-390X

Research Article

Keywords: ecological niche property; medicine–food homology; property predilection; research protocol; blinded assessment; quality ecology

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11786216/v1

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

Additional Declarations: The authors declare no competing interests.

A “One Species, Two Standards” Bidirectional Regulation Protocol Based on Ecological Niche Property Encoding

Guanfeng Yang

Abstract

The production side of Chinese medicine has long run two parallel systems of objectives: medicinal-material production aims to make medicinal properties pronounced, whereas food-material production aims to keep the directional predilection balanced. The two, however, have lacked a unified formal objective function and a testable research design. Drawing on the ecological niche property (ENP) framework—distinct from the suitable-area modeling of the ecological niche model (ENM); in this paper “niche” denotes a habitat coding niche—this study defines habitat parameters as the coding variables of medicinal properties and the whole medicinal-property vector, which connects directly to formulation rules, as the output variable. It proposes a “sender-free encoding” model of rule-governed mapping and, on that basis, designs a symmetric, parallel bidirectional regulation experiment on the same species: a medicinal line that strengthens the predilection and a food line that attenuates it. The primary endpoint is the medicinal-property vector measured by a two-layer measurement instrument (Instrument A: multi-source literature agreement; Instrument B: a blinded expert panel). Three hypothesis lines, E1–E3, are set for reliability, construct validity, and between-group effect size, with falsification conditions and sample-size re-estimation rules specified in advance. The paper is organized by evidence status. Part I reports the completed calibration of Instrument A: 10 species and 6 sources, with the four mainstream sources showing perfect agreement (1.000) across three dimensions. After the extended sources were included, four-nature $\kappa = 0.913$ and Gwet AC1 = 0.927; for meridian tropism, pairwise set agreement was 0.813, species-level complete agreement 0.600, and subset agreement 1.000, with all meridian disagreements being strict subset additions; E1 received exploratory support. A sensitivity analysis treating “not recorded” as a valid category yielded $\kappa = 0.424$, indicating that the complete-case framing is optimistic. Part II presents the preregistered content for E2 and E3: the primary effect size is fixed as Cliff’s δ ($\delta \geq 0.33$), the pilot includes no fewer than 12 units per line, and the confirmatory sample size is approximately 64 units per group. This protocol does not promise clinical efficacy; it tests only the measurability and regulability of the medicinal-property construct.

Keywords

ecological niche property; medicine–food homology; property predilection; research protocol; blinded assessment; quality ecology

1 Introduction

1.1 Background and problem statement

The production practice of Chinese medicine has long faced two opposite objectives at once. As a medicinal material, production pursues the distinctiveness and stability of nature, flavor, and action; as a food-medicine substance, it pursues a balanced predilection suitable for long-term daily consumption. The same species requires opposite quality directions on two production lines, yet existing quality systems are organized mainly around constituent content, geographic reputation, and standardized production, lacking a unified objective function that takes the “desired medicinal property” directly as output and allows habitat parameters to be solved inversely.

1.2 Conceptual reversal: from noise to signal

In traditional pharmacognosy, environmentally induced trait variation has long been classified as "environmental modification"; in discussions of the formation mechanism of daodi materials, environmental variation is largely treated as a source of variation to be controlled and explained. This study proposes a different positioning: medicinal property is not a quality disturbed by the environment but information encoded by habitat. Habitat parameters—temperature, light, water, soil, food source, symbiosis, stress, and so on—follow a testable, rule-governed mapping and, through the species' physiological processes, are translated into a medicinal-property vector with direction and magnitude.

This reversal does not attribute intentionality to the environment. The encoding referred to here is sender-free encoding, a rule-governed mapping that presupposes no intentional agent. The Shannon engineering model itself requires only the formal structure of a signal source, channel, and receiver, and does not require the sender to possess intentionality. This position is compatible with discussions of non-living sign processes in biosemiotics and niche construction theory; the paper treats it as a general theoretical background rather than claiming a new philosophical proposition.

1.3 Demarcation from prior work

To avoid a blanket claim of priority, the objective functions of this study and seven types of prior work are compared in Table 1.

Table 1 Comparison of objective functions (demarcation)

Work category	Representative path	Objective function	Relationship to this study
Quality ecology; environmental-modification theory	Wan Deguang et al.	Explanation of quality variation and "two-way interaction"	Both recognize the environment's role; this study sets habitat explicitly as a coding variable and outputs a property vector
Stress-effect theory; habitat-simulating cultivation	Huang Luqi; Guo Lanping	Simulating native stress and enhancing secondary metabolism	Practical predecessor of the medicinal line (strengthening); its reverse attenuation has no systematic scheme and is supplied by the food line
Geographic ecological-suitability regionalization	Chen Shilin; GMPGIS	Cross-region site selection and ecological similarity	Solves "where to plant"; this study is the two-line differentiation within one producing site
Ecological niche model	MaxEnt suitable-area prediction	Habitat suitability and suitable area	Answers "where it can be grown" without outputting a property vector
Controlled environment and stress regulation	Elicitation; chronoculture	Secondary-metabolite constituent content	Optimizes constituents rather than the whole property; not connected to formulation rules
Standardized production	Chinese-medicinal-material GAP	Standardized production quality	Provides an institutional framework without an inverse objective function
General sign theory	Biosemiotics; niche construction	Sign processes and ecological information	Theoretical support; this study is its concrete application in the property domain

It should be emphasized that this study does not claim the above lines of work cannot guide production; rather, it points out that none of their objective functions is a whole medicinal-property vector directly connected to formulation rules. The increment of this study lies in making the objective function explicit and equipping it with testable measurement and experimental design.

1.4 Significance and structure

Theoretically, this study provides a unified formal language for the "habitat–property–formulation" chain; practically, it offers an engineering scheme for the medicinal and food lines of the same species, with opposite directions but shared parameters; methodologically, it discloses hypotheses, indicators, and falsification conditions in advance so that subsequent empirical work is reproducible. Section 2 presents the theoretical framework, coding model, and a formulation worked example; Section 3 gives the version freeze record, hypotheses, and measurement instruments; from Section 4 onward, the experimental design, statistical plan, falsification conditions, safety and compliance, submission strategy, and discussion are presented in order.

2 Theoretical framework

2.1 Construct definition

Ecological niche property (ENP) denotes the overall characteristic of medicinal property formed by rule-governed mapping from habitat parameters under the constraint of the species' genetic background, operationalized as a medicinal-property vector with direction and magnitude. The "medicinal property" referred to here is the operationalized expression of traditional constructs such as the four natures, five flavors, and meridian tropism; it is not equivalent to clinical efficacy, and efficacy propositions are not within the scope of this protocol. The medicinal or edible part is recorded separately as an additional dimension of the vector: the part derives from developmental processes rather than habitat parameters, so it does not enter the independent variables of the habitat coding function f, but it participates in determining ascending–descending–floating–sinking.

2.2 Formal notation

Let the habitat parameter vector be $x = (x1, \ldots, xk)$, corresponding respectively to observable, recordable ecological variables such as temperature, light, water, soil, food source, and stress. Let the coding rule f be the E/M/G three-dimensional scoring mapping: E for the Environment dimension, M for the Material dimension, and G for the Genesis dimension, with weights and scales specified in the frozen scoring manual. The medicinal-property vector is $y = f(x)$, whose components include direction (e.g., positive or negative on the cold–hot axis) and magnitude (discrete levels and positions on a continuous spectrum), with an attached part dimension d. Predilection coordinates are expressed as $0/+1/-1$ for balance and the two opposite directions, while the continuous spectrum describes the degree of deviation.

On the production side, the objective function can be written as approximation to a target medicinal-property vector y^*: the medicinal line selects the combination of habitat parameters that increases predilection magnitude, whereas the food line selects the combination that converges the predilection toward the zero point. The two lines share the same parameter space and have opposite target directions; this is the formal statement of "one species, two standards." It should be noted that ecological parameters are high-dimensional and continuous whereas property outputs are discrete and ordinal, so inverse solving is underdetermined and admits multiple solutions; this study does not claim the inverse problem has been made well-posed, but rather reduces its ill-posedness through an explicit objective function and constraints.

2.3 Six elements of the coding model

Table 2 Elements of the coding model

Communication-model element	Corresponding entity in this study
Sender	Habitat parameters (a signal source without intention)
Coding rule	The E/M/G mapping f

Channel and carrier	Species physiological processes: metabolism, tissues, stress response
Message	The property vector y: direction × magnitude (with part dimension)
Receiver and decoding	Formulation rules: three-component formulation, harmony index, usage levels
Noise	Label measurement error: contemporaneous multi-source disagreement and judge disagreement
Validity check	Blind re-review, out-of-fold validation, and intervention experiments

2.4 Operational definition of noise

The model defines noise as the measurement error of property labels, comprising two distinguishable sources: contemporaneous multi-source framing disagreement and expert judgment disagreement. The magnitude of noise is estimated by blinded-agreement coefficients (the dual framing of κ and Gwet AC1) and calibrated by the blind re-review procedure. The dual framing is chosen to mitigate the instability of κ when prevalence is extreme. It should be noted that the E/M/G weights, magnitude scales, and part dimension are all frozen parameters, and noise is the only free parameter explicitly registered through a measurement procedure; its estimation method is registered rather than left to post hoc explanation.

2.5 Measurement and evidence hierarchy

Whether the whole medicinal-property vector can serve as a measurable construct depends on whether evidence is delivered step by step: first measurement reliability (blinded agreement), then content validity (frozen procedures and expert review), then construct validity (out-of-fold prediction and known-group discrimination), and finally predictive validity (long-term clinical or nutritional comparison, which this protocol does not promise).

2.6 Worked example: from the property vector to formulation position (method demonstration)

So that "once computed, it can directly be used in formulation" does not remain abstract, the labels already confirmed by Instrument A are used as an example to show how the property vector directly enters the position rules of the three-component formulation (heaven-ascending · earth-anchoring · water-sinking, in the ratio 50/25/25), with the whole process not passing through any constituent-content conversion.

Table 3 Demonstration from the property vector to formulation position (ascending/descending judged by the three-element criterion)

Species–part	Latin name	Property vector y (direction × magnitude; flavor; meridian)	Ascending/descending judgment (nature–flavor + part + action)	Formulation position
Perilla leaf	Perilla frutescens (L.) Britt.	Warm × regular; pungent; lung, spleen	Pungent-warm + leaf + release the exterior and disperse cold; mainly ascending and dispersing	Heaven-ascending position (leading; lifting and exterior-releasing)
Perilla seed	Perilla frutescens (L.) Britt.	Warm × regular; pungent; lung	Pungent-warm + seed + direct qi downward and resolve phlegm; mainly descending and sinking	Water-sinking position (directing qi downward)

Cassia seed	Senna obtusifolia (L.) H.S.Irwin & Barneby	Cold × slight; sweet-bitter-salty; liver, large intestine	Cold-cool + seed + clear heat and promote bowel movement; mainly descending and sinking	Water-sinking position (clear heat; purge downward)
Purslane (whole herb)	Portulaca oleracea L.	Cold × regular; sour; liver, large intestine	Sour-cold + whole herb + clear heat and cool blood; mainly descending and sinking	Water-sinking position (clear heat; cool blood)

Operational meaning: ascending–descending–floating–sinking is not determined by nature and flavor alone; the frozen manual stipulates a three-element criterion of "nature–flavor + part + action." Once the property vector y is obtained from the coding rule f, its role and proportion in a formula can be determined directly by consulting the position rules of the three-component formulation according to ascending/descending, the five flavors, and meridian tropism, after which the harmony index checks whether the whole is balanced—without first measuring a constituent and then passing through multiple rounds of conversion. It must be stated that this table demonstrates a method process, not a clinical prescription, and does not constitute any efficacy promise.

3 Version freeze record, hypotheses, and measurement instruments

3.1 Version and freeze record

To distinguish "ex ante freezing" from "post hoc reading," the freezing order of the key files is recorded in Table 4 (①②③ indicate order; absolute times are based on the preprint registration timestamp in §7.1). All thresholds were frozen once with SOP v1.0 and were not changed during sample expansion.

Table 4 Version and freeze record

Freeze order	File/version	Time (2026)	Event
①	SOP v1.0	November 16	Judgment fields, inclusion rules, and the E1–E3 threshold framing were frozen once
②	First pilot batch	November 16	Four species closed; κ/AC1 = 1.000
③	Expanded-sample data	From November 16	Six species added and a six-source matrix built; the frozen thresholds were unchanged

3.2 Hypotheses

Based on the formal framework in Section 2, three experimental hypothesis lines, E1–E3, are set, fully decoupled from H1–H10 in prior preregistration texts to avoid cross-reference confusion.

E1 (reliability hypothesis; exploratory support): under the frozen judgment procedure, the judgments of the property vector from each source show agreement above chance.

E2 (construct validity hypothesis; preregistered): the forward model based on habitat parameters predicts the property vector out-of-fold significantly better than the prespecified baseline.

E3 (regulation effect hypothesis; preregistered): the medicinal and food lines show a significant between-group difference in the blinded property vector.

3.3 Frozen thresholds and interpretation framing

Following general interpretation standards in the published literature, thresholds were frozen with SOP v1.0 (Table 5); they must not be adjusted post hoc during analysis. It should be noted that E1 at the current 10-species stage reports only exploratory reliability, and formal interpretation must wait until the number of items reaches 30 or more.

Table 5 Hypothesis thresholds and literature basis

Hypothesis	Indicator	Frozen threshold	Literature basis
E1	κ; Gwet AC1	Point estimate ≥ 0.81; lower CI ≥ 0.61 (formal interpretation when n ≥ 30)	Landis & Koch (1977); Gwet (2008)
E2	Out-of-fold AUC	Point estimate ≥ 0.70; lower CI > 0.50	Swets (1988); baseline is the chance level 0.5
E3	Cliff's δ (ordinal primary indicator)	δ ≥ 0.33; CI does not cover 0	Cliff (1993); δ ≥ 0.33 corresponds to a medium effect
Sample size	Pilot units / confirmatory sample	Pilot ≥ 12; confirmatory about 64 units per group	Julious (2005); Cohen (1988)

Plain-language meaning of the thresholds: above 0.81 counts as "almost perfect agreement," and 0.61–0.80 as "substantial agreement"; an AUC of 0.5 is equivalent to guessing, and only above 0.7 is there practical discrimination; Cliff's δ measures the proportion by which one group is stronger than the other, with 0.33 being the medium-difference threshold; preparing 12 units per group for the pilot is only a feasibility starting point, and to detect a medium difference the confirmatory experiment requires about 64 units per group.

3.4 Two-layer design of the measurement instrument

Given the constraints of the research conditions, the "predilection measuring instrument" is designed as a two-layer structure that can be implemented in stages: Instrument A is multi-source literature agreement, accomplishable by desk work alone without a laboratory; Instrument B is a blinded expert panel, launched when cooperation conditions are available. The two layers share the same set of judgment fields and analysis code, and the results of Instrument B can be directly merged with those of Instrument A for verification.

Table 6 Comparison of the two-layer measurement instruments

Item	Instrument A: multi-source literature agreement	Instrument B: blinded expert panel
Judges	Six layered sources (see Table 7)	Three to five licensed TCM physicians or identification specialists
Implementation conditions	Public materials and desk work; no laboratory needed	Launched after cooperation conditions become available
Blinding requirements	Double entry; independent source extraction	Blinded to group and to regulation direction
Measurement target	Inter-source reproducibility	The property vector of regulated samples
Analysis indicators	κ, Gwet AC1, and confidence intervals	κ, Gwet AC1, and Cliff's δ

3.5 Instrument A: multi-source literature agreement (Part I: completed calibration)

The six judgment sources are layered by credibility, and this paper does not describe them as mutually independent judges: sources may share upstream origins and transcription relationships, so the agreement indices measure source reproducibility without correcting for common-source dependence. The source layering is shown in Table 7.

Table 7 Judgment source layering

Level	Source	Role in the analysis
L1 State/textbook	Chinese Pharmacopoeia (2025 edition); the state-planned textbook Chinese Materia Medica	Mainstream framing; the agreement conclusions are based on this
L3 Professional platform	Xiaohe Medical Dictionary; the China Pharmaceutical Information Query Platform (dayi)	A professional supplement to the mainstream framing
L4 Institutional outreach / encyclopedia	The medicinal botanical garden of Nanjing University of Chinese Medicine; Douyin Encyclopedia	Stress-test sources verifying the instrument's detection ability

Current progress: after SOP v1.0 was frozen, on the basis of the first batch of four high-consensus species (perilla leaf, perilla seed, cassia seed, and purslane), six species were added—coix seed, goji berry, chrysanthemum, adzuki bean, poria, and donkey-hide gelatin—for a total of 10 species and 6 sources. The results show that the four mainstream sources were in complete agreement on the three dimensions of four natures, five flavors, and meridian tropism, with κ and Gwet AC1 both equal to 1.000. After the two extended sources were included, four-nature raw-category κ = 0.913 and AC1 = 0.927 (95% CIs 0.712–1.000 and 0.780–1.000, respectively), while four-nature direction coding remained 1.000; the five flavors were in complete agreement; for meridian tropism, per-meridian κ was 0.788–1.000 and AC1 0.905–1.000, with pairwise set agreement 0.813, species-level complete agreement 0.600, and subset agreement 1.000. All disagreements can be located to specific cells: category-boundary differences such as coix seed "cool / slightly cold," and single-meridian additions in meridian tropism for perilla leaf, perilla seed, cassia seed, and goji berry, all of which are strict subset relationships. κ = 1 proves only reliability, not validity.

Sensitivity analysis: recomputing with "not recorded" treated as a valid category (rather than removed as missing) yielded four-nature κ = 0.424, markedly lower than the 0.913 of the complete-case framing, indicating that the pairwise complete-data framing has an optimistic bias. The extended sources have structural non-recording for chrysanthemum, adzuki bean, poria, donkey-hide gelatin, and others; therefore 0.913 reflects the reproducibility of entries recorded by all sources and cannot represent the overall coverage quality of the extended sources. This difference is disclosed in the limitations.

3.6 Instrument B: blinded expert panel (Part II: to be launched)

When cooperation conditions are available, three to five licensed TCM physicians or Chinese-materia-medica identification specialists are engaged to form a blinded panel that independently judges the regulated samples; the judges are unaware of group allocation and regulation direction, and sample numbering and physical delivery are handled by non-judging personnel.

The frozen SOP should stipulate judge admission and training, the trial-rating agreement threshold (per the E1 threshold in Table 5), rating forms and judgment levels, disagreement arbitration rules, and withdrawal and replacement rules. The methodological anchors are sensory-evaluation panels, scale-development standards (including COSMIN), and the Delphi expert-consensus method.

3.7 Measurement endpoints and judgment fields

The primary endpoint is the blinded property vector (direction and magnitude), appearing as a source judgment matrix in Instrument A and an expert judgment matrix in Instrument B; the secondary endpoints are constituent content and safety indicators, obtained from public literature or cooperative testing, and do not replace the primary endpoint.

Table 8 Judgment field list

Field	Description

Record number	A unique identifier corresponding one-to-one with the original records
Species name and botanical origin	Including the Latin name and origin verification
Part	The medicinal or edible part, recorded separately and participating in ascending/descending judgment
Source or judge number	Instrument A records the source; Instrument B records an anonymous number
Property direction and magnitude	Direction judgment and magnitude level, filled per the frozen framing
Judgment basis	An original-text excerpt and the source location
Entry and review information	Entry person, reviewer, and date

3.8 Data sources for E2

E2 does not depend on future experiments; its data foundation is the already constructed habitat–property paired dataset (the unified master table, containing objective habitat parameters and property labels), and out-of-fold validation is carried out on that dataset according to the preregistered number of folds. The definitions of habitat parameters and the framing of property labels are both frozen in SOP v1.0, avoiding variables being set after data reading.

3.9 Division of labor and execution conditions

Documentation and computation are completed independently by the author using reproducible scripts and public data, covering all SOPs, forms, code, data management, statistical analysis, sample-size simulation, and feasibility argumentation; the items requiring personnel or partners are limited to expert time, field or base implementation, and testing channels. The lowest-cost real-data channel is Instrument A, whose implementation does not depend on a laboratory or dedicated funding but only on accessible public materials and reproducible processes.

4 Experimental design

4.1 Species selection criteria

Before viewing any regulation results, the pilot species are frozen according to four criteria: first, being listed in the current medicine–food homology directory, so both production lines have regulatory legitimacy; second, being annual or short-cycle, so that a first readout is obtainable within one natural year and the three-to-five-year cycles of perennial woody plants do not stall the protocol; third, having public literature reporting associations between habitat factors and the species' nature–flavor or related constituents, so the regulation direction has ex ante evidence rather than subjective expectation; and fourth, having a mature cultivation basis and public agronomic materials, so habitat parameters are recordable and reproducible.

By these criteria, annual herbs such as perilla (including leaf and seed) and purslane are priority candidates; the final pilot species and their grouping are confirmed in the registration text, and this manuscript does not presuppose unverified yield or effect figures.

4.2 Symmetric, parallel bidirectional design

The same species is deployed on two production lines simultaneously, with the same parameter space and opposite target directions. The medicinal line aims to strengthen the predilection of the target vector y*, regulating temperature, light, water, soil, stress intensity, harvest time, and other parameters so that the direction remains unchanged while magnitude increases; the food line aims to converge the predilection toward zero, selecting within the same parameter space combinations that weaken magnitude and move toward neutrality. The two lines are implemented in the same period and share a control baseline, forming the symmetric structure of "one species, two standards."

It is worth pointing out where the novelty of this symmetric structure lies. The predilection-strengthening direction of the medicinal line is continuous with the stress-effect theory and habitat-simulating cultivation of Huang Luqi and Guo Lanping: simulating native stress to promote secondary metabolism already has mature practice, and this study replaces its target from constituent content with the whole property vector directly connected to formulation rules. What truly lacks systematic precedent is the predilection-attenuating direction of the food line: removing stress in reverse within the same parameter space so that the predilection converges toward balance is a symmetric target equal to strengthening but long absent. Chinese medicine is not without evidence of the same substance having two natures. The Shennong's Classic of Materia Medica classically records ginseng as slightly cold, whereas the current Chinese Pharmacopoeia records ginseng as slightly warm and red ginseng as warm; this historical and caliber difference shows that the same species can present two property directions. After ginseng is steamed into red ginseng, the processed product is suited to those with yang deficiency and weak qi and blood. Such a change occurs at the post-harvest processing stage, whereas this protocol moves it to the pre-harvest cultivation end, the two belonging to successive stages that can be layered rather than substituted.

Table 9 Design comparison of the two production lines

Design element	Medicinal line (strengthening)	Food line (attenuating)
Target direction	Predilection magnitude increases while direction remains unchanged	Predilection converges toward the zero point
Parameter space	The same set of habitat parameters	The same set of habitat parameters
Objective function	Approximation to a strengthened y*	Approximation to a neutral y*
Primary endpoint	Blinded property vector	Blinded property vector
Regulatory positioning	Medicinal-material quality management	Food-medicine substance; reduced predilection exposure

4.3 Regulation factors and recording requirements

Candidate regulation factors include observable, recordable habitat variables: temperature and its accumulation, light duration and intensity, water supply and drought stress, soil type and fertility, biotic or abiotic stress, symbiotic plant/animal conditions, harvest time, and processing method. Each treatment unit must completely record factor levels and implementation times, ensuring external reproducibility from the records; the correspondence between factors and the E/M/G three-dimensional scores of the coding rule f is specified in the frozen scoring manual.

4.4 Randomization, numbering, and blinding

Treatment units are allocated to the two lines and factor levels according to randomization rules; sample numbers are generated by personnel not participating in judgment, and grouping and regulation direction are kept confidential from Instrument B judges throughout. Physical delivery, rating-form collection, and data locking are separated in the division of labor, so judges can see only numbered samples and cannot infer treatment conditions.

4.5 Internal pilot and sample-size re-estimation

A nested internal pilot is set before the confirmatory experiment, with each production line first preparing no fewer than 12 units (Table 5): after the first round of regulation, a small-sample blinded readout is performed, and based on its variance and effect-size estimates the confirmatory sample size is re-determined according to formulas stipulated in advance within the protocol. Estimated for a medium effect at power 0.80, the confirmatory sample size is about 64 units per group, corrected by the pilot variance. This re-estimation is a prescribed action within the registration text, used only for sample-size

calculation and not to modify hypotheses, endpoints, or thresholds; without pilot support, the confirmatory experiment proceeds at a conservative sample size or is paused and disclosed.

The conversion between threshold and power: under the normal approximation, Cliff's $\delta = 0.33$ corresponds to Cohen's $d \approx 0.60$; 64 units per group at two-tailed $\alpha = 0.05$ and power 0.80 can detect $d \approx 0.50$, so the power for the threshold effect is above 0.80 and the sample-size criterion is conservative for $\delta \geq 0.33$.

5 Statistical analysis plan

5.1 General principles

The statistical plan is frozen before data locking, and the main analysis addresses only the preregistered primary endpoints and hypotheses; all analyses provide runnable code and reproducible data, with key results presented in dual framing or through cross-checks.

5.2 E1 reliability analysis

For single-category dimensions (such as four-nature direction), Fleiss κ and Gwet AC1 are computed, with pairwise Cohen κ as a cross-check; for multi-label dimensions (five flavors and meridian tropism), binary calculations are first made per category, and pairwise set agreement is then reported. Confidence intervals use bootstrap resampling over items: with the ten items as resampling units (rather than source pairs), 2,000 resamples are drawn and the 2.5%/97.5% percentiles taken; judgments follow the Table 5 thresholds, and when $n < 30$ only exploratory reliability is reported, with the resolution limits disclosed.

5.3 E2 construct validity analysis

On the constructed habitat–property paired dataset, a forward model is fitted with habitat parameters as input and the property vector as output, using out-of-fold validation (leave-one-out or a prespecified number of folds) to assess prediction. For categorical framings, AUC and related discrimination indicators are reported and compared with the prespecified baseline model (AUC = 0.5); E2 is supported only when the out-of-fold AUC point estimate is ≥ 0.70 and the lower confidence bound is above 0.50. Multi-label and ordinal outputs are handled per the frozen framing, without post hoc parameter tuning to improve results.

5.4 E3 regulation effect analysis

The blinded property vectors of the two lines serve as the between-group comparison object, with Cliff's δ (the ordinal primary indicator) computed per dimension and its confidence interval reported, supplemented by vector-direction agreement. Multiple comparisons are corrected by the Benjamini–Hochberg (BH) method, and E3 is supported only when the δ point estimate is ≥ 0.33 and the confidence interval does not cover 0. Missing data are analyzed as pairwise complete cases per the SOP with the missing mechanism disclosed, without unprescribed imputation.

5.5 Software, recomputation, and missing handling

All computations use publicly available statistical software and independent scripts, saving both intermediate matrices and final results to support line-by-line recomputation; source absence is recorded as "not recorded," the cell is removed in the main analysis, and the "not recorded as a category" sensitivity analysis (currently $\kappa = 0.424$) tests framing robustness, with filling from other sources prohibited. Any anomalies found during analysis are recorded only as revision items entering the next version, without altering already frozen framings.

6 Falsification conditions and safety compliance

6.1 Prespecified falsification conditions

To make the protocol refutable by data, three falsification lines and their consequences are prescribed in advance in Table 10.

Table 10 Falsification conditions and consequences

Hypothesis	Falsification condition	Consequence
E1 reliability	Agreement falls below the frozen Table 5 threshold or is not better than chance	Revise the judgment SOP and postpone Instrument B and the experiments
E2 construct validity	Out-of-fold AUC is not better than the prespecified baseline	Abandon predictive claims for the forward model while retaining the framework discussion
E3 regulation effect	The two lines' blinded vectors show no significant between-group difference	Do not claim regulability; report the negative result and conduct a review

Falsification results are reported truthfully as negative or neutral findings, not avoided by post hoc endpoint changes, threshold adjustments, or selective reporting.

6.2 Safety argument for the food line

For substances in the medicine–food homology directory, the food line takes "balancing" as its goal, directionally reducing predilection exposure and the predilection dose in long-term daily consumption, consistent with food regulation's goals of balance, safety, and long-term intake. The candidate factors (temperature, light, water, stress, and so on) are all cultivation environment parameters and involve no exogenous additives; nevertheless, the net effect of each factor on safety must be confirmed by testing, and safety indicators are monitored concurrently as secondary endpoints.

6.3 Compliance boundaries and expression discipline

This protocol promises only the measurement and regulation of the property construct, not disease treatment or clinical efficacy; external texts must not contain efficacy claims such as treating or curing, and food-related expressions comply with advertising and food-safety regulation. Regarding existing schools, it avoids asserting that they "cannot do it," and instead states that "the objective function is not a whole property vector"; it does not claim the inverse problem is "well-posed," but only that it "reduces ill-posedness"; entropy-related expressions distinguish the measurement, analogy, and rhetorical levels.

6.4 Data ethics and traceability

Expert and partner information is managed under informed consent and the minimum-necessary principle; original records, numbering mappings, judgment bases, and version changes are documented throughout. The registration text, SOPs, data, and scripts are archived uniformly, with every revision marked by version and reason, ensuring the whole process is traceable and reproducible.

7 Submission strategy and schedule

7.1 Dual-track submission

Given that Chinese-medicine journals vary in their familiarity with the protocol genre, a dual-track strategy is adopted: the formal track goes to journals that maintain "protocol/design" columns or accept registered reports; concurrently, the work is first registered on a preprint platform to obtain a verifiable timestamp. The timestamp is essential for establishing priority, whereas journal acceptance is incremental, and the two do not replace each other. Part I already contains the Instrument A calibration

results, which can reduce the risk of rejection for “no data”; the specific journals and platforms are confirmed before submission.

7.2 Paper sequence and dependencies

The writing order is arranged by evidence dependency: the protocol paper and the “three-factor adaptation operationalization” nine-grid matrix paper do not depend on field experiments and can be drafted in parallel; Instrument A has completed the SOP freezing and the measurement of two batches totaling 10 species; Instrument B and the field experiments launch when cooperation conditions are available; the ENP forward-model empirical paper depends on validity evidence and is placed after it; the pilot paper involving nutritional or clinical comparison is arranged last.

Table 11 Paper sequence and preconditions

Order	Product	Precondition	Current status
1	The “one species, two standards” protocol paper	Hypotheses and processes frozen	This manuscript
1	The nine-grid matrix paper (theoretical discussion)	Does not depend on experimental data	Completed in the same period
2	The Instrument A result report	SOP v1.0 and sample expansion	Completed (supporting data)
3	The ENP forward-model empirical paper	Instrument B and validity evidence	Awaiting external conditions
4	The nutritional/clinical comparison pilot paper	Comparison design and cooperation channels	Arranged last

7.3 Scheduling principles

The desk work (documentation, SOP, scripts, Instrument A, and the matrix paper) proceeds continuously under the current conditions, with no nodes dependent on external funding; steps requiring expert time, bases, or testing channels are triggered when the conditions become available, and all prefabricable materials are completed before the trigger, avoiding waiting for internal preparation after external resources arrive.

8 Discussion and limitations

8.1 Expected contributions

If the hypotheses are supported under the prespecified conditions, the defensible contribution is to take the whole property vector, directly connected to formulation rules, as the objective function; to formalize the “habitat–property–formulation” chain; and to support bidirectional regulation of the medicinal and food lines of the same species. Its differences from quality ecology, suitable-area models, stress regulation, and GAP are demarcated through objective functions rather than established by negating opponents.

8.2 Main limitations

First, the current 10 species of Instrument A remain a small sample dominated by common directory substances, so E1 is only exploratory reliability and the results cannot be extrapolated to the whole directory; animal-derived and highly contested varieties have not yet been included. Second, Instrument B has not been launched, so the reliability of the blinded panel and the actual readouts of regulated samples remain blank; current evidence stays at the literature-reliability level, and multi-source agreement proves only inter-source reproducibility, not that the labels are true or valid. Third, the complete-case framing has an optimistic bias: the sensitivity analysis (“not recorded” as a category) gives $\kappa = 0.424$, and its gap from 0.913 shows that the missing structure affects agreement estimation.

Fourth, under the research conditions there is no laboratory or dedicated funding, so constituent and safety data can only rely on public literature or future cooperative testing; the validity of the property labels themselves has not been independently verified, the inverse problem remains underdetermined with multiple solutions, and this protocol only reduces the ill-posedness of the inverse problem without rendering it well-posed.

8.3 Honest disclosure and genre arrangement

The paper is divided by evidence status: Part I truthfully reports the completed calibration of Instrument A, with E1 marked as exploratory support; Part II keeps E2 and E3 preregistered, disclosing only the hypotheses, processes, and falsification conditions without reporting conclusions in advance. This arrangement both matches the current evidence strength and explains the necessity of the protocol genre, avoiding packaging an unverified framework as an empirical result.

8.4 Future work

Subsequent work proceeds in the order: continue expanding the Instrument A sample to 30 species and include highly contested varieties; launch the Instrument B blinded assessment and freeze the thresholds; complete the annual pilot regulation experiment; and then re-evaluate the forward model and the empirical paper. The coding rules for animal-derived food materials, the food-line safety data, and the validation of a broader range of ecological parameters are listed separately as framework extensions and are not promised in advance within this protocol.

References

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics, 1977, 33(1): 159–174.

Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. British Journal of Mathematical and Statistical Psychology, 2008, 61(1): 29–48.

Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceutical Statistics, 2005, 4(4): 287–291.

Swets JA. Measuring the accuracy of diagnostic systems. Science, 1988, 240(4857): 1285–1293.

Cliff N. Dominance statistics: ordinal analyses to answer ordinal questions. Psychological Bulletin, 1993, 114(3): 494–509.

Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum Associates, 1988.

Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons, 1981.

Chinese Pharmacopoeia Commission. Pharmacopoeia of the People’s Republic of China (Volume I). 2025 ed. Beijing: China Medical Science Press, 2025.

Huang LF, Yan Q, Wu CY, et al. From the environmental-modification theory of Chinese medicine quality to Chinese medicine quality ecology: inheritance and innovation of Professor Wan Deguang’s academic thought. Journal of Chengdu University of Traditional Chinese Medicine, 2026, 49(3): 14–20.

Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. Ecological Modelling, 2006, 190(3–4): 231–259.

Dsouza A, Dixon M, Shukla M, Graham T. Harnessing controlled-environment systems for enhanced production of medicinal plants. Journal of Experimental Botany, 2025, 76(1): 76–93.

Basso LF, Baluška F, Souza GM. Plant agency: from senome and electrome to plant semiosphere. Frontiers in Plant Physiology, 2026, 4: 1861112.

Odling-Smee FJ, Laland KN, Feldman MW. Niche construction: the neglected process in evolution. Princeton: Princeton University Press, 2003.

Wang Q. Constitutional Theory of Traditional Chinese Medicine. Beijing: People’s Medical Publishing House, 2005.

Lipsitz LA, Goldberger AL. Loss of “complexity” and aging: potential applications of fractals and chaos theory to senescence. JAMA, 1992, 267(13): 1806–1809.

National Medical Products Administration; Ministry of Agriculture and Rural Affairs; National Forestry and Grassland Administration; National Administration of Traditional Chinese Medicine. Announcement on issuing the Good Agricultural Practice for Chinese Medicinal Materials (No. 22 of 2022). 2022-03-17.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 1995, 57(1): 289–300.

Shennong Ben Cao Jing (Shennong’s Classic of Materia Medica). Ed. Sun Xingyan, Sun Fengyi. Beijing: Zhonghua Book Company, 1985.

Guo LP, Zhou LY, Kang CZ, et al. Strategies of medicinal plants adapting to environmental stress and the “habitat-simulating cultivation” of daodi medicinal materials. China Journal of Chinese Materia Medica, 2020, 45(9): 1969–1974.

Meng XX, Huang LF, Dong LL, et al. Global producing-region ecological suitability and quality ecology of Panax notoginseng. Acta Pharmaceutica Sinica, 2016, 51(9): 1483–1493.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- InstrumentAExtendedResultsWorkbook.xlsx