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.

Computational Formalization of the Principle of Correcting Bias with Counter-Bias: A Same-Coordinate Per-Axis Balancing Model with Full-Scale Testing of Classical Formulas

Guanfeng Yang

jjaarr@163.com

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

Research Article

Keywords: correcting bias with counter-bias, computational pattern differentiation, same coordinate, per-axis balancing, classical formulas, incompatibility contraindication, auditable artificial intelligence

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11856472/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.

Computational Formalization of the Principle of Correcting Bias with Counter-Bias: A Same-Coordinate Per-Axis Balancing Model with Full-Scale Testing of Classical Formulas

Yang Guanfeng

Abstract: Correcting bias with counter-bias is a clearly stateable principle of traditional Chinese medicine (TCM) treatment based on pattern differentiation, yet it has long existed in empirical language and lacks a reproducible, falsifiable formalization. Rather than quantifying the entire yin–yang philosophy, this paper narrows the defensible boundary to a computational formalization of this principle family. The subject state and candidate formula are placed on the same set of measurable coordinates, with zero as the origin. Along four dimensions—cold–heat, ascending–descending, supplementing–draining, and meridian tropism—signed vectors with magnitude are computed axis by axis, and the residual and balancing determine whether a formula matches, how large the deviation is, and whether it is safe to use; the explanation is produced as a direct by-product of the decision process. Using a closed corpus of 246 independent classical formulas (100 from the Shanghan Lun and 146 from the Jinkui Yaolue) and 158 herbs, full-scale testing is conducted under two dose-conversion schemes. All formulas are 100% computable with zero missing values. Under real doses, 97.8% of warm formulas have a positive net value and 93.8% of cold formulas have a negative net value; the two schemes show 99.2% sign agreement and a net-value rank correlation of 0.998. The claim that formulas are generally balanced to zero is falsified: fewer than one quarter have an absolute net value below one, and net values form a continuous spectrum rather than natural three bands. Ascending–descending is an independent axis, with only moderate correlation at the herb level and sign divergence in 20.3% of formulas at the formula level. Supplementing–draining and meridian tropism are weakly correlated with the directional axes and serve directional and localizing functions, respectively. The two-level contraindication gate triggers on only four formulas (1.63%), all of which are classical opposing-yet-complementary precedents handled by mandatory review and whitelist grading rather than a blanket veto. On the formula-generation task, case retrieval plus modification raises coverage from 13.8% for from-scratch generation to 72.0%, with 98.2% average coverage at the efficacy-label layer. This paper offers a

reproducible, auditable, falsifiable method-layer model and its computational evidence on a classical corpus; it does not claim clinical efficacy or causal mechanisms. The system is positioned as an assistive tool for physician review: it does not automatically generate prescriptions or replace physicians, and clinical validity requires separate formal studies.

Keywords: correcting bias with counter-bias; computational pattern differentiation; same coordinate; per-axis balancing; classical formulas; incompatibility contraindication; auditable artificial intelligence

1 Introduction

1.1 Problem statement

Treatment based on pattern differentiation is the core process of TCM clinical practice, and its operational point is commonly summarized in statements such as “when there is cold, apply heat; when there is heat, apply cold; when there is deficiency, supplement; when there is excess, drain.” These statements share one structure: first identify the bias of the subject state on a given dimension, then apply an intervention in the opposite direction to correct it, returning the system to relative balance. This paper collectively refers to this principle family as “correcting bias with counter-bias.” Its logical structure is not mysterious, yet it has long been carried in empirical language, its transmission relies on individual insight, and its results cannot be independently reproduced or clearly falsified.

This creates two practical problems. First, at the level of knowledge form, the principle is a set of explicitly stateable rules without a corresponding mathematical expression, so there is no stable way to check what result should follow from the rules. Second, at the engineering and regulatory level, when artificial intelligence enters TCM assistive settings, black-box probabilistic outputs can neither explain the basis of each decision nor satisfy the requirements of accountability and human review. If a principle can itself be rule-based, writing it as a deterministic, reproducible form is more fundamental than merely approximating its conclusions with a statistical model.

1.2 Research objectives and the upper bound of claims

The objective is to provide a computational formalization of the principle of correcting bias with counter-bias and to conduct full-scale testing on a clearly bounded closed corpus, answering four research questions: (RQ1) Can the principle be written as deterministic per-

axis computation such that every formula in the corpus is computable with no missing values; (RQ2) Does the computed direction agree with the classical annotated cold–heat direction, and is the conclusion robust across dose-conversion schemes; (RQ3) Do cold–heat, ascending–descending, supplementing–draining, and meridian tropism constitute independent dimensions with distinct responsibilities rather than repetitions of the same information; (RQ4) On downstream tasks such as contraindication gating and formula generation, can the formalization produce checkable, auditable results.

The boundary of the claims must be set in advance. This paper does not claim to quantify the entire yin–yang philosophy as numbers, nor does it assert that computed results equal clinical efficacy or reveal biological mechanisms. The defensible statement is limited to: this paper provides a computational formalization of the principle family and demonstrates that it is reproducible, falsifiable, and process-auditable on 246 classical formulas. Correlation is not causation, and directional agreement in offline computation does not automatically imply clinical effectiveness; the model's ability to extrapolate beyond the corpus also requires separate validation.

1.3 Research approach and contributions

The research approach is “same coordinate, origin, per-axis, balancing.” The subject state and candidate formula are placed in the same set of measurable coordinates, with zero as the origin representing balance. Measurement of the subject yields a state with direction and intensity, and computation of the formula yields a signed vector with magnitude. Residuals are computed axis by axis, and balancing is judged when the two cancel on that axis. Formulas themselves carry sign and magnitude; demanding that every formula be balanced to zero is an error, and the duty of the system is to report residuals and risks faithfully rather than output a preset answer.

There are four contributions. First, a same-coordinate balancing model with zero as the origin and four-axis computation is proposed, with explicit operators and conventions. Second, full-scale computability with zero missing values is achieved on 246 classical formulas, together with dual-scheme robustness testing. Third, through correlation and sign-divergence analysis, the independence and division of labor among dimensions are empirically established, and the early hypothesis that formulas are generally balanced to zero is thereby falsified. Fourth, the two-level contraindication gate and retrieval-based formula generation are incorporated into the same deterministic framework, so every decision can be logged and reviewed.

1.4 Organization of the paper

Section 2 reviews related work and positions the study; Section 3 presents the theoretical framework and formal definitions; Section 4 explains the data, dose conversion, and research methods; Section 5 reports the full-scale test results; Section 6 discusses the findings, methodological significance, limitations, and compliance boundaries; Section 7 gives the conclusions and future research directions.

2 Related Work and Positioning

2.1 The quantitative lineage of TCM attributes

Quantifying the cold–heat attributes of Chinese medicines is not a new topic. Existing studies follow different paths: some map the four qi onto ordered levels based on expert scores and construct cold–heat indices; some use relative dosage as a clue to infer attribute intensity from the proportions of medicines in a formula; others introduce modern physicochemical indicators and attempt correlation analysis of medicinal properties using composition or biological activity. These works provide a basis for ordering attributes, but most remain at the single-dimensional measurement or explanation level, and few treat "measurement–balancing–decision–gating" as a closed computational process.

The difference here is that the work does not stop at scoring medicines. Instead, attribute ordinals are inserted into the same coordinates, cancellable signed operators are defined, and the research object is advanced from "whether a given herb is cold or hot" to "for a given subject and candidate formula, how large the per-axis deviation is, whether it is balanced, and whether it can be used," with judgments made for the entire corpus rather than illustrated by examples.

2.2 Artificial intelligence and knowledge graphs in TCM

In recent years, knowledge graphs, rule engines, and large language models have been used for TCM-assisted pattern differentiation, recommendation, and question answering. Knowledge-driven systems ensure controllability through structured rules but tend to have rigid rules and remove cold–heat hedging in a one-size-fits-all manner. Large models offer natural-language interaction but suffer from unstable conclusions, untraceable bases, and poor auditability. Newly emerging TCM multi-agent retrieval-augmented schemes have begun to emphasize rule bases and logging, consistent with the auditability orientation here,

but they are usually dominated by fixed-band risk rules and lack a unified coordinate origin and per-axis residual computation.

The present model is positioned as a deterministic computation engine rather than a conversational agent: all conclusions are derived directly from explicit rules without relying on black-box probabilities, so every output can be traced back to a specific axis, weight, and rule. It can serve as an explainable backend for upper-layer question answering or large-model systems rather than a competing chat interface.

2.3 Differentiation from prior research

Overall, the differentiation lies not in data scale or interaction form but in three points: first, the same-coordinate design with zero as the origin enables direct signed operations between subject and formula; second, the multiple dimensions are treated as independent axes, acknowledging and preserving the classical structure of opposing yet complementary elements rather than treating all hedging as error; third, both propositions and conclusions are reproducible and falsifiable, supported by full-scale statistics rather than representative cases. Section 6 further explains these differences against specific systems.

3 Theoretical Framework and Formalization

3.1 Same coordinate and origin

The first cornerstone of the model is one set of coordinates and a clear origin. Zero represents balance, meaning no identifiable directional bias on that dimension. Both measurement of the subject state and computation of the formula fall on the same coordinates, so the two can be directly added and subtracted without empirical conversion between two scales. This design turns the judgment of whether a subject and formula match into a residual problem within the coordinates.

Let the subject vector be P and the formula vector F, each composed of the same set of dimensions. The core of the decision is to compute the residual of $D = P + F$ on each axis: when the residual on an axis is zero, that axis is balanced; when it is nonzero, its sign indicates which side remains biased and its absolute value indicates the degree of deviation. It should be emphasized that the goal is to compute per-axis residuals faithfully rather than forcibly rewrite every formula to zero; the sign and magnitude of a formula carry information in themselves.

3.2 The four dimensions and their responsibilities

The model is built on four dimensions. The cold–heat axis captures the warming–cooling direction, the most direct expression of correcting bias with counter-bias. The ascending–descending axis captures the ascending and descending movement of qi and is a directional dimension independent of cold–heat. The supplementing–draining axis captures the tendency toward consolidation/supplementation versus purgation/draining. The meridian axis captures the organ-level localization of formula action and is itself a directionless, weight-conserving localization vector. The first three axes carry direction and can be positive or negative; the fourth handles localization. Their responsibilities differ, and this is tested through correlation and divergence counts in Sections 5.3 to 5.6.

Table 1 Four-qi ordinals and seven-level ascending–descending (quantitative conventions)

Four qi	Great heat	Heat	Warm	Slight warm	Neutral	Slight cold/cool	Cold	Great cold
Ordinal LV	10	7	4	2	0	−2	−5	−10
Ascending–descending	Ascend	Slight ascend	Micro-ascend	Neutral	Micro-descend	Slight descend	Descend	
Coefficient S	+1	+0.5	+0.3	0	−0.3	−0.5	−1	

The four-qi ordinal maps eight levels of medicinal properties onto asymmetric ordered values with zero centered. Ascending–descending uses seven levels, from +1 for ascending to −1 for descending, with zero as neutral. Among the 158 herbs in the corpus, the fine ascending–descending bands (±0.3) cover 43 herbs, about 27%, and are ordinary rather than rare exceptions; the model therefore retains the full seven levels rather than simplifying to five.

3.3 Balancing operators and net value

On the cold–heat axis, let a formula consist of k herbs, with the four-qi ordinal of herb i being LV_i and its composite weight w_i. The cold–heat net value and harmony degree are defined as: net = Σ(w_i·LV_i) / Σw_i; C = |Σ(w_i·LV_i)| / Σ(w_i·|LV_i|). The net value gives the signed direction and intensity of the formula on the cold–heat axis. The harmony degree C lies between zero and one and measures the degree of same-direction medicinal properties within the formula: the closer C is to one, the more consistent the direction; the

closer to zero, the more intense the internal cold–heat hedging. The ascending net value snet and ascending harmony degree cs are computed in the same form by replacing LV with S.

The composite weight w_i is synthesized from two parts. First, block weights reflect the relative importance of different dimensions in the overall decision: cold–heat, five flavors, meridian, and ascending–descending take 0.40, 0.25, 0.25, and 0.10, respectively. Second, sovereign–minister–assistant–courier role weights take 3, 2, 1.5, and 1, reflecting formula structure. The two sets of weights are merged after being determined in their respective stages, with explicit rules that are not adjusted according to results.

3.4 Computation of supplementing–draining and meridian

The supplementing–draining axis is computed by two layers of deterministic rules. First, the four qi classify warming/heat and cold/cool. Then the supplementing–draining tendencies of the five flavors are added: sweet, sour, and astringent mainly consolidate/supplement, while bitter, acrid, salty, and bland mainly drain, with the bland flavor having the strongest draining tendency; containing sweet-warm strengthens supplementation and sweet-cold weakens it, and the final result is clipped to the interval [−1, 1]. The meridian axis distributes the dosage weight of each herb equally among the meridians it enters, and after aggregation over all herbs forms a weight-conserving localization vector on the twelve meridians without assigning directional signs.

3.5 Generation of decision states and explanations

On top of the four-axis computation, the model gives a three-state decision for each axis: balanced, positive residual, or negative residual, with an intensity band according to the absolute residual. The overall conclusion is formed by combining the states of the axes and outputs risk flags together with the contraindication gate. The explanation is not separately written text afterward but a direct restatement of the decision process: it cites the residual signs and magnitudes on specific axes, the triggered rules, and the whitelist basis, so the explanation and conclusion necessarily agree and can be checked item by item. This is the key distinction of a decision model relative to a pure explanation model.

4 Data and Research Methods

4.1 Closed corpus and unit of analysis

The study uses a closed corpus with the formula as the unit of analysis. The corpus contains 246 independent classical formulas, including 100 from the Shanghan Lun and 146 from the Jinkui Yaolue, involving 158 herbs. As verified by code, all formulas originate solely from the two classical works, without later-era or modern empirical formulas. The benefit of a closed corpus is clear boundaries and item-by-item checkability; the limitation is the limited extrapolation range of conclusions, addressed specifically in Section 6.

Each formula records its name, source, function, expected direction, composition, and herb weights. The expected direction is annotated according to classical function as warm, cold, or neutral/harmonizing, with 91, 48, and 107 formulas respectively among the 246. This annotation serves as the reference for direction testing and does not participate in computation, avoiding the use of conclusions to test conclusions.

4.2 Dose conversion and dual schemes

Classical dosage units differ from modern metric units, so the study builds an explicit dose-conversion table and uses two schemes for cross-validation. Schemes A and B use different conversion coefficients for weight units; for example, one “liang” converts to about 13.80 g and 15.625 g, respectively. Drugs given by volume, count, or special utensils are handled by their own conversion rules, such as banxia, jingmi, mangxiao, dazao, and fuzi listed separately, and pills do not participate in weight conversion.

In addition, a prior scheme P that does not rely on real doses is constructed: equal or role weights replace real doses, using only drug attributes. The three schemes coexist, allowing us to distinguish whether conclusions are determined by drug attributes or by specific doses and to test the sensitivity of sign conclusions to the conversion method.

4.3 Computation flow

The complete flow is: parse formula composition and doses, convert by scheme to obtain herb weights; query the four qi, five flavors, meridians, and ascending–descending herb by herb and compute supplementing–draining values; aggregate on the four axes to obtain net values, harmony degrees, and localization vectors; run the contraindication gate and mark triggers; combine axis states to generate decisions and explanations. The whole

process is deterministic rules, without random optimization or black-box models, and the same inputs always yield the same outputs.

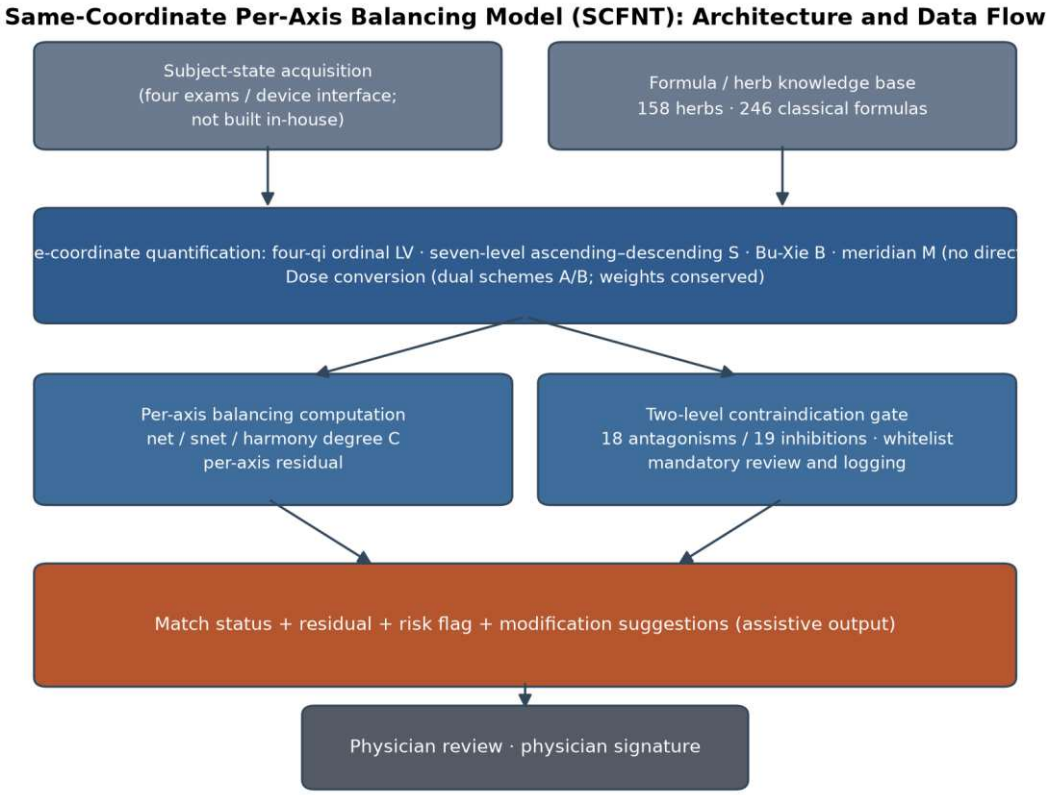

Fig. 1 Overall architecture and data flow of the same-coordinate per-axis balancing model

4.4 Two-level contraindication gate

The gate handles incompatibilities in two levels. The first level scans contraindication pairs such as the eighteen antagonisms and nineteen mutual inhibitions at the formula layer. Once triggered, it does not directly veto but enters the second level of mandatory review: if the combination belongs to a classical precedent of opposing yet complementary use, it is graded and logged for release according to a whitelist; if it is not a precedent, it is blocked with a prompt. The gate output fully records the triggered pair, rule source, and handling basis to meet audit requirements. Specific triggers are given in Section 5.8.

4.5 Formula generation and evaluation conventions

Formula generation compares two paths: from-scratch automatic generation selects within the herb library by greedy or two-stage residual, while case retrieval plus modification first retrieves the closest classical formula as a template and then guides a small number of additions by residual. Evaluation focuses on coverage (the proportion of targets met), Jaccard herb overlap, and residual, supplemented at the efficacy-label layer and

retrieval precision. Efficacy labels map the 158 herbs to 20 efficacy classes, with primary and secondary classes taking 1 and 0.5. Retrieval precision is measured by P@1, P@5, and same-family hits within the same function class and compared with a random baseline.

The symptom-to-pathomechanism evaluation is divided into two subsets: a small sample is constructed for formulas with clean symptom descriptions and clear attribution, and full-scale evaluation is conducted for all formulas from which symptoms can be extracted, reported as the Top-1 hit rate, with its position explicitly defined as assistance for physician verification rather than automatic diagnosis.

5 Results

5.1 Full-scale computability and zero missing values (RQ1)

Running the computation flow on the 246 formulas one by one, all formulas obtain valid results on the four axes, for a 100% computability rate with no missing values and no interruptions caused by data gaps. For occasional non-common medicines in the original text that have no corresponding attributes within this dataset, the rules explicitly skip that herb and mark it in the log, without affecting the remaining herbs or the whole-formula computation. This shows that the per-axis formalization of correcting bias with counter-bias is closed on the closed corpus: every formula can be deterministically computed rather than holding only for representative cases.

5.2 Cold–heat direction concordance and robustness (RQ2)

Using the classical annotated expected direction as the reference, under real-dose scheme A, 89 of the 91 warm formulas have a positive net value, a 97.8% direction concordance; 45 of the 48 cold formulas have a negative net value, a 93.8% concordance. The computed direction is highly consistent with classical function and is derived directly from drug attributes without using annotation information.

Table 2 Cold–heat direction concordance and dual-scheme robustness

Test item	N	Result	Concordance/correlation
Warm formulas	91	Positive net 89	97.8%
Cold formulas	48	Negative net 45	93.8%
A/B sign agreement	240	Agree 238	99.2%
A/B net-value ranking	246	Spearman 0.998	—
Prior P / real A ranking	246	Spearman 0.943	—

Under the two dose-conversion schemes, the cold–heat sign agreement for nonzero formulas is 99.2% (238/240), and the Spearman correlation of net-value rankings reaches 0.998, indicating that directional conclusions are insensitive to the dose-conversion method. More notably, the rank correlation between the prior scheme P, which contains no real doses, and the real-dose scheme A still reaches 0.943, indicating that the cold–heat direction of a formula is mainly determined by the attributes and composition of the drugs themselves, while specific doses mainly affect intensity rather than direction.

5.3 Falsification of the "balanced-to-zero" hypothesis

A natural conjecture is that since classical formulas emphasize balance, they are generally balanced to zero on the cold–heat axis. The data give a negative answer. Formulas with an absolute net value below one account for only 19.9%, 24.0%, and 23.2% under the P, A, and B schemes respectively, about one fifth of the total; the vast majority retain a clear direction. Further examining the net-value distribution, net forms a continuous spectrum rather than three separated bands, so the "three bands" do not exist naturally, and the two cut points 1.54 and 3.07 are drawn artificially for convenient grading.

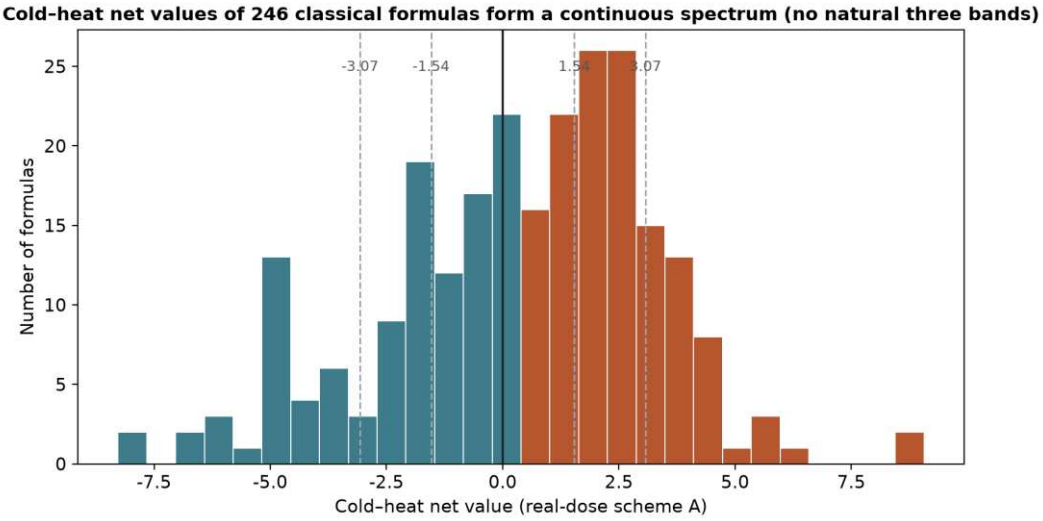

Fig. 2 The cold–heat net values of the 246 classical formulas form a continuous spectrum

The harmony degree C portrays the within-formula structure from the other side. Under real-dose scheme A, 51 formulas have C below 0.3, 43 fall between 0.3 and 0.6, and 152 are no lower than 0.6. Most formulas have same-direction medicinal properties and a high C, while about one fifth show marked hedging, suggesting that classical formulas do not avoid deliberate combined use of cold and warm. Accordingly, the early hypothesis that formulas

generally return to zero is falsified by statistical evidence and abandoned; the model no longer targets returning to zero but faithfully retains the sign and magnitude of formulas.

5.4 Independence of the ascending–descending axis (RQ3)

At the herb level, the Pearson correlation between the four-qi ordinal LV and the ascending–descending coefficient S for the 158 herbs is 0.4391 and the Spearman correlation is 0.5019, only moderate; 25 herbs show sign divergence in cold–heat versus ascending direction, meaning the cold–heat bias and ascending direction are opposite. This shows that heat does not equal ascending and cold does not equal descending, and ascending cannot be inferred from cold–heat.

At the formula level, the Spearman correlation between the two axes is about 0.67 and the Pearson about 0.69, still short of being mutually substitutable. Under the two-decimal display convention, 50 formulas show sign divergence on the two axes, 20.3%; under full-precision strict opposite-sign judgment it is 52, the difference coming from two formulas with net values near zero (Mahuang Lianqiao Chixiaodou Tang and Zhizhu San (蜘蛛散)), whose ascending net values are zero at display precision. To avoid treating numerical noise as direction, the main text uses the 50-formula, 20.3% convention consistent with the workbook.

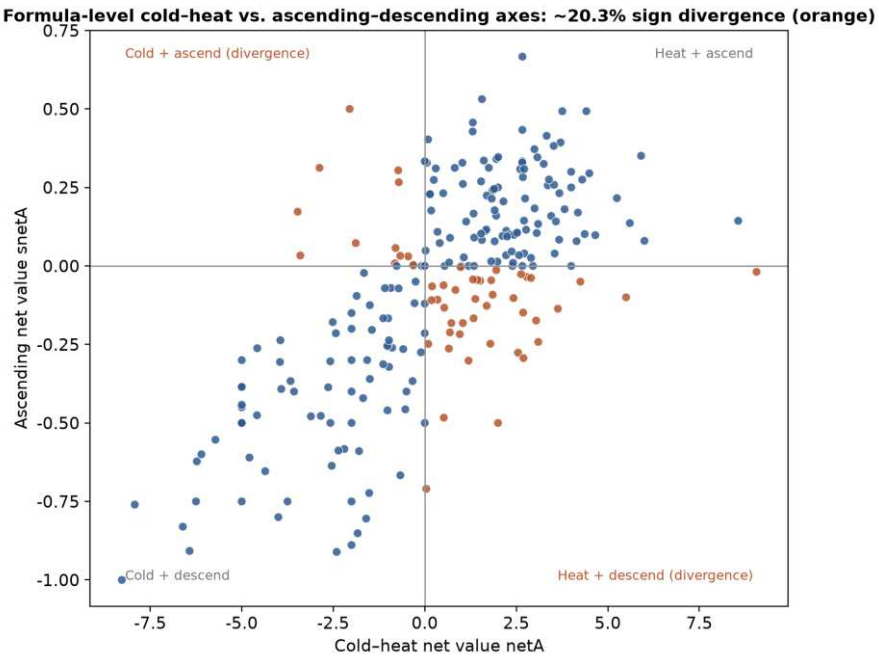

Fig. 3 Sign divergence between the formula-level cold–heat and ascending–descending axes

In addition, 11 formulas have an ascending net value exactly zero at full precision, reflecting a neutral structure different from the cold–heat axis. The herb-level and formula-

level evidence together support that ascending–descending is an independent qi-movement axis and should be computed separately rather than as an appendage of cold–heat.

5.5 The supplementing–draining axis and three-axis fingerprints

The supplementing–draining axis shows only weak correlation with the other two directional axes: its correlation with the cold–heat net value is about 0.29 and with the ascending net value about 0.32, indicating that it carries independent tendency information rather than repeating cold–heat or ascending. Threshold judgment is strongly affected by the convention, so the convention is stated explicitly: at a threshold of 0.1 the counts are 75, 44, and 127; at 0.2 they are 59, 80, and 107. The main text uses the 0.2 convention with balanced three-band counts.

Table 3 Three-axis fingerprints and direction counts

Item	Count	Item	Count
Theoretical sign combinations	27	Observed 25	2 unobserved
Number of directional axes 0	11	1 axis: 61	2 axes: 98
Number of directional axes 3	76	With ≥1 axis: 235	95.5%

Combining the three directional axes, there are 27 theoretical sign combinations, and 25 are observed in the 246 formulas, showing diverse axis structures. The distribution of the number of directional axes is 11 formulas with zero, 61 with one, 98 with two, and 76 with three; 235 formulas, 95.5%, have at least one directional axis. This further contradicts the “generally balanced-to-zero” conjecture and supports the rationality of per-axis computation.

5.6 The meridian axis and near-orthogonal localization

The meridian axis is a directionless localization vector. Its correlations with the three directional axes are all very weak: −0.108 with the cold–heat net value, −0.039 with the ascending net value, and −0.008 with supplementing–draining. The correlation between localization entropy and the cold–heat net value is only 0.049, and the largest absolute value of the relevant correlations is about 0.11, close to orthogonal. This confirms that meridian localization does not overlap with directional axes and should be handled separately as a localization dimension.

The primary meridians of the 246 formulas are concentrated in the middle jiao and lung: spleen 82, lung 68, stomach 39, heart 21, liver 17, and large intestine 8. The

distribution broadly matches the clinical picture of the two classical works, which emphasize middle-jiao regulation and exterior-release and lung systems, and the localization vector conserves weight, making it suitable for organ-level targeting rather than directional balancing.

5.7 Formula generation and retrieval (RQ4)

From-scratch generation performs poorly. Within the 158-herb library, the coverage of target composition is only 13.8%, Jaccard overlap 9.4%, and residual 0.111, indicating that pure combinatorial generation is hard to restore classical formulas. Case retrieval plus modification is markedly better: on the 196 formulas with clear efficacy labels, coverage reaches 72.0%, Jaccard 52.9%, residual 0.093, with an average of 1.13 added herbs and a template-found rate of 54.6%, roughly a 5.2-fold improvement over from-scratch coverage.

Table 4 Formula generation and retrieval evaluation

Path	Coverage	Jaccard	Residual	Avg. additions
From scratch	13.8%	9.4%	0.111	—
Retrieval + modification	72.0%	52.9%	0.093	1.13
Efficacy-label layer	98.2%	≥99.9% for 189	—	—

At the efficacy-label layer, the 196 formulas have an average coverage of 98.2%, and 189 reach no lower than 99.9%, indicating that the model is highly accurate in restoring efficacy even when exact herbs differ. On retrieval precision, within the 246-formula library, P@1 rises from 0.557 to 0.577 and P@5 from 0.463 to 0.502, and same-family hits within the top five rise from 0.683 to 0.756, all well above the random baseline of 0.158.

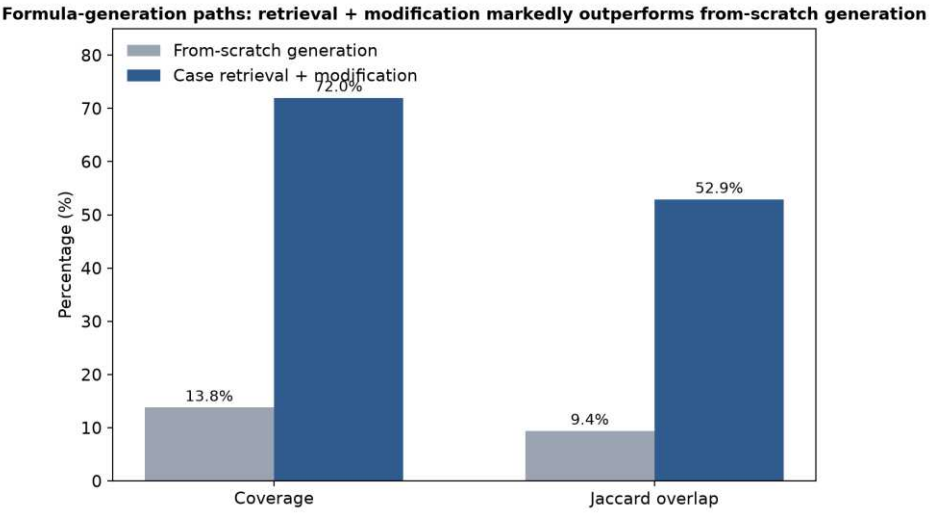

Fig. 4 Coverage comparison of the two generation paths

5.8 Two-level contraindication gate results

The first-level scan finds four formulas triggering the eighteen antagonisms: Fuzi Jingmi Tang, Chi Wan, Gansui Banxia Tang, and Zhuye Tang, 1.63% of the 246. All four are classical precedents of opposing yet complementary use, so they are graded and released according to the whitelist at the second level, with full records of the triggered pair and basis. By source, the 100 Shanghan formulas have zero triggers, while the 146 Jinkui formulas have four. The nineteen mutual inhibitions have zero triggers, because the relevant herbs are not in the 158-herb dataset, and this structural uncheckability is honestly marked rather than claimed as none.

The gate results show that the classical texts do not absolutely avoid contraindication combinations, but they appear only in a small number of specific contexts and are classical precedents. The two-level design avoids both indiscriminate veto and indiscriminate release, and the whitelist plus logs makes each handling auditable.

5.9 Symptom-to-pathomechanism prediction

Symptoms can be extracted for 185 of the 246 formulas, and the remaining 61 lack sufficiently clear symptom descriptions and are not included. On the clean subset of 22 formulas with clear attribution, the Top-1 hit rate is 21/22 = 95.5%; on the full 185 formulas the Top-1 hit rate is 63/185 = 34.1%. The sharp drop indicates that full-scale symptom reasoning is affected by label granularity and text complexity, and the clean subset reflects the ceiling under simple conditions.

The two miss cases are classic label-granularity problems rather than model errors. Xiaoqinglong Tang targets “exterior release” but the model predicts “phlegm-resolving, cough-suppressing, and panting-calming,” and Banxia Xiexin Tang is labeled “exterior release” although its function is centered on harmonizing the middle jiao. Both involve overlapping function labels of different granularity. Given that the model is positioned as assistance for physician verification, such cases should be judged by the physician rather than automatically overruled.

5.10 Case study: Sini Tang

Sini Tang is a representative formula for rescuing yang, and its full-scale and case-level results are consistent. The cold–heat net values under the three schemes are all positive: 6.778 for P, 4.659 for A, and 4.507 for B, with the harmony degree C equal to one under all

schemes, indicating completely consistent warming properties and no internal hedging. The ascending net values are about 0.1: 0.100 for P, 0.098 for A, and 0.101 for B, a gentle ascending tendency.

Table 5 Sini Tang multi-scheme results

Indicator	Scheme P	Scheme A	Scheme B
Cold–heat net value	6.778	4.659	4.507
Harmony degree C	1.000	1.000	1.000
Ascending net value	0.100	0.098	0.101

The case shows that the three schemes agree on direction and harmony and differ only in intensity, consistent with the full-scale conclusion that direction is determined by attributes and doses affect intensity. The model's explanation for Sini Tang directly cites the strong positive cold–heat residual and full harmony degree, supporting the yang-rescuing position with computable quantities rather than rhetorical description.

5.11 Limitations and reliability

The study has several limitations. First, the closed corpus is limited to the two classical works, and extrapolation to later formulas and modern empirical formulas requires further validation. Second, the dose conversion relies on specific historical conventions, and although dual schemes and a prior scheme show sign stability, the exact intensity remains convention-dependent. Third, efficacy and symptom labels carry granularity problems, and the full-scale symptom Top-1 hit rate is only 34.1%.

On reliability, the computations are deterministic and reproducible, and the accompanying data package provides per-table SHA-256 fingerprints and a full fingerprint over seven tables, with a documented serialization scheme so that third parties can independently recompute and verify whether data were modified after release. The current annotation reliability is based on rule-based plus single-author cross-checking; a second independent human annotator is not yet in place and is explicitly marked as a follow-up rather than overstated.

6 Discussion

6.1 From explanation to decision

Existing quantification work largely explains existing formulas: after the formula is given, it computes vectors and similarity to interpret why the formula works. The model here

goes a step further toward a decision model. Given a subject state and candidate formulas, it computes per-axis residuals and outputs balanced, positive, or negative states with intensity and risk flags, and the explanation is a direct restatement of the residual process. The full-scale 95.5% directional-axis rate and high direction concordance show that the model does not merely describe but can deterministically judge whether a formula fits and how large the residual is.

6.2 Why not force balance to zero

The falsification of the balanced-to-zero hypothesis has theoretical implications. Classical formulas are not generally neutralized; about four fifths retain a clear direction, and a substantial subset has high same-direction harmony. This is consistent with the logic of correcting bias with counter-bias: a formula deliberately carries a directional bias to counter the subject's opposite bias, and the two together approach the origin. Forcing the formula itself to zero would erase the therapeutic force. The model therefore treats the formula as a signed vector with magnitude and judges balance only after combining subject and formula, which is more faithful to the principle than treating zero as an intrinsic property of the formula.

6.3 Independence and near-orthogonality of the axes

The correlations and divergence counts support the multi-axis design. Ascending–descending is only moderately correlated with cold–heat and has 25 divergent herbs and 20.3% divergent formulas; supplementing–draining is weakly correlated with the first two; meridian is nearly orthogonal to all directional axes. Collapsing these into one score would lose movement, consolidation versus drainage, and localization information. Per-axis computation preserves the structure and allows the explanation to point to the specific axis and rule responsible, which is the basis for auditability and physician verification.

6.4 Positioning and boundaries under regulation

The model is a computer-assisted decision support tool rather than an autonomous prescribing system. Under Chinese regulation, prescriptions must be issued by the treating physician, and AI may not automatically generate prescriptions or replace the physician. The design respects this boundary: the gate and symptom modules output suggestions and flags for human review, with the whitelist and logs retaining a complete trail. This positioning is also what keeps the relevant inventions within patentable subject matter as auxiliary devices

and data-processing methods, rather than methods of disease diagnosis or treatment. Clinical adoption therefore depends on integration into physician workflows and validation on real-world cohorts, which are the next steps.

7 Conclusion

This paper formalizes the classical Chinese medicine principle of correcting bias with counter-bias into a deterministic same-coordinate per-axis balancing model. On the same coordinates, cold–heat, ascending–descending, supplementing–draining, and meridian are computed separately, and the residual of subject plus formula on each axis determines whether the combination is balanced, the sign and magnitude of any deviation, and the risk flags. The full-scale test on 246 classical formulas from the Shanghan Lun and Jinkui Yaolue, involving 158 herbs, yields four main findings: all formulas are computable with zero missing values; cold–heat direction concordance is about 94–98%, robust across dual dose schemes and a prior scheme; ascending–descending, supplementing–draining, and meridian are demonstrably independent or nearly orthogonal axes, with 95.5% of formulas carrying at least one directional axis; and retrieval plus modification clearly outperforms from-scratch generation, while the two-level gate and audit logs keep the process safe and reviewable.

The balanced-to-zero hypothesis is falsified by the data, and the model instead treats formulas as signed vectors with magnitude and positions itself as physician-facing decision support. Future work will expand the corpus beyond the two classical works, introduce a second independent human annotator for reliability, and validate the model on real-world clinical cohorts, while retaining the reproducible, auditable, and falsifiable character that a computational account of correcting bias with counter-bias requires.

References

[1] Huangdi Neijing, Suwen (Treatise on the Supreme Truth in the Plain Questions), classical TCM text.

[2] Zhang Zhongjing. Shanghan Lun (Treatise on Cold Damage) [M]. Eastern Han dynasty, critical edition.

[3] Zhang Zhongjing. Jinkui Yaolue (Essential Prescriptions from the Golden Cabinet) [M]. Eastern Han dynasty, critical edition.

[4] National Health Commission of the People's Republic of China, National Administration of Traditional Chinese Medicine. Administrative Rules for Internet Diagnosis and Treatment (Trial) [Z]. 2022.
[5] National Medical Products Administration. Implementation Opinions on “Artificial Intelligence plus Drug Regulation” [Z]. 2026.
[6] National Medical Products Administration. Key Information Table of Famous Ancient Classical Formulas (43 formulas) [Z]. Public document.
[7] Zhao Xinyu, Li Jiawei, Kan Hongxing. Research on a TCM Diagnosis and Treatment Agent System Based on Multi-Agent Collaboration and Retrieval Augmentation [J]. Journal of Nanjing University of Chinese Medicine, 2026, 42(11): 1754-1769.
[8] TCM-RAG-Agent. Open-source Project of a TCM Rule Engine with Retrieval Augmentation [CP/OL]. 2026.
[9] Chinese Journal of Experimental Traditional Medical Formulae. Modern Pharmacological and Clinical Studies on Sini Tang and Its Modified Formulas [J]. 2026.
[10] Ant Group. Official Press Release on the Positioning of the Health Assistant “Afu” [Z]. 2026.

Note: This is a preprint draft. The edition and page numbers of classical texts and the full bibliographic details of modern references will be completed according to the target journal's style before formal submission; public links to the regulatory documents and platform materials are provided in the appendix for verification.

Appendix A Data and Code Availability

The research corpus, attribute tables, dose-conversion tables, and all computation code are implemented as deterministic rules, and given the same inputs they reproduce every number in this paper. The per-axis net values, harmony degrees, gate triggers, and formula-generation results for the 246 formulas are stored as checkable workbooks and data files. Formal statements on data availability, authorship, and usage licenses will be completed according to the requirements of the chosen preprint platform.

Appendix B Quick Reference of Key Conventions

Four-qi ordinals: great heat 10, heat 7, warm 4, slight warm 2, neutral 0, slight cold/cool −2, cold −5, great cold −10. Seven-level ascending–descending: +1, +0.5, +0.3, 0, −0.3, −0.5, −1. Net value: net = $\Sigma(w\cdot LV)/\Sigma w$; harmony degree C = $|\Sigma(w\cdot LV)|/\Sigma(w\cdot|LV|)$; the ascending net value and harmony degree are computed in the same form by replacing LV

with S. Block weights: cold–heat 0.40, five flavors 0.25, meridian 0.25, ascending–descending 0.10; sovereign–minister–assistant–courier role weights 3/2/1.5/1.

Direction counts: warm 91, cold 48, neutral/harmonizing 107; gate triggers on four formulas (1.63%); formula-generation coverage 13.8% to 72.0%; efficacy coverage 98.2%; symptom Top-1 95.5% on the clean subset and 34.1% on the full set. All percentages use 246 formulas (or the corresponding subsample) as the denominator, with conventions consistent with the research workbooks.

Supplementary Files

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

- SCFNTDataPackageandVerifiableArchive.xlsx