ScieRxiv
NOT PEER-REVIEWED
Original Article

# A population–individual reference interval conversion model for personalized hypertension management (without altering universal guideline thresholds): A real-world cohort study in rural primary care

**Xinjuan Zhang[1], Junwei Song[2], Yong Feng[3], Xiaoyun Yang[4], Jianmin Yao[1], Yu Xu[5], Jialan Yan[5], Wenxiu Liu[6], Lianyuan Feng[5,*]**

[1]Department of Neurology, Nanhe District People's Hospital, Xingtai 054400, Hebei, China.
[2]Department of Orthopedics, Nanhe District People's Hospital, Xingtai 054400, Hebei, China.
[3]Department of Blood Transfusion, Bethune International Peace Hospital, Shijiazhuang 050082, Hebei, China.
[4]Department of Public Health, Nanhe District People's Hospital, Xingtai 054400, Hebei, China.
[5]Department of Neurology, Bethune International Peace Hospital, Shijiazhuang 050082, Hebei, China.
[6]Department of Cardiology, Bethune International Peace Hospital, Shijiazhuang 050082, Hebei, China.

***Correspondence to:** Dr. Lianyuan Feng, Department of Neurology, Bethune International Peace Hospital, Shijiazhuang 050082, Hebei, China. E-mail: fengly2005@163.com

**Received: 19 November 2026 | Approved: 24 November 2026 | Online: 24 November 2026**

## Abstract

**Aim:** Guidelines apply one blood pressure (BP) threshold to every patient. We asked whether a personal steady-state reference band (PSR-band) derived from a patient's

© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

own serial readings improves discrimination of clinically judged BP status beyond the measured reading, without altering that threshold.

**Methods:** In a rural primary-care cohort of 200 patients with hypertension (79 men; mean age 66.1 years), the band was derived from visits 1–14 and evaluated on the held-out visit 15. Visit-15 readings and their corrected counterparts were compared with patient tolerance and physician judgment by dual-anchor ROC analysis with paired DeLong tests and Cohen's kappa. A score of 5 defined the positive class.

**Results:** Measured readings discriminated strongly: diastolic BP (DBP) against patient tolerance (AUC 0.829) and systolic BP (SBP) against physician judgment (AUC 0.901). Correction did not improve either index—ΔAUC −0.017 for DBP ($P$ = 0.556) and +0.006 for SBP ($P$ = 0.871) against patient tolerance—and for SBP against physician judgment it was significantly worse (0.843 *vs.* 0.901; $P$ = 0.010). Grading agreement was poor (kappa 0.112–0.307). No band coincided with the population interval (0/200) and none was fully disjoint (0/200); 73.0% of systolic bands partially overlapped it and 26.5% of systolic readings were classified discordantly.

**Conclusion:** A guideline-compliant PSR-band adds a second description of the same reading but does not sharpen discrimination, and may blunt it for systolic pressure. It is an interpretive layer, not a better classifier.

**Keywords:** Hypertension, personalized medicine, reference interval conversion, blood pressure management, precision medicine

**Abbreviations**

ABPM: ambulatory blood pressure monitoring

AUC: area under the curve

BP: blood pressure

CI: confidence interval

CV: coefficient of variation

DBP: diastolic blood pressure

DBPc: corrected diastolic blood pressure

GEE: generalized estimating equations

HR: heart rate

HRc: corrected heart rate

ICC: intraclass correlation coefficient

IRB: institutional review board

LMM: linear mixed-effects model

PSR-band: personal steady-state reference band

REML: restricted maximum likelihood

ROC: receiver operating characteristic

SBP: systolic blood pressure

SBPc: corrected systolic blood pressure

SD: standard deviation

## 1. INTRODUCTION

Hypertension is the leading modifiable risk factor for cardiovascular and cerebrovascular disease worldwide, affecting approximately 1.28 billion adults. In China, the number of patients with hypertension has reached 245 million, imposing a substantial burden on the health care system, particularly in rural primary care[1,2]. Current international and national hypertension guidelines establish uniform diagnostic thresholds and treatment targets derived from large-scale epidemiological evidence, which standardizes care and facilitates widespread implementation[1,3,4]. A single threshold nevertheless conceals the within-person variability that determines how representative any one reading is[22]. However, this population-based, one-size-fits-all approach inherently overlooks inter-individual differences in physiological baseline, comorbidity profile, and treatment tolerance. The same absolute BP value may carry vastly different clinical implications for a patient with a lifelong low baseline than for one with a chronically high baseline, potentially leading to either excessive intervention or insufficient treatment[5,6]. The heavy burden of hypertension in rural primary care[7] and the established association between elevated BP and cerebrovascular events[8] underscore the need for assessment tools that take each patient's own physiological steady state into account.

To address this gap, we developed a population-individual reference interval conversion model based on a patented algorithm[9,10]. Its core innovation is distributional mapping: a patient's percentile rank in his or her own derivation distribution is mapped to the corresponding percentile rank of the population reference distribution, without altering any universal guideline threshold. We validated the model in a real-world rural cohort of patients with hypertension.

We systematically compared population-based standard assessment with the individualized corrected assessment in terms of grading consistency, baseline adjustment for sex, and temporal stability across serial follow-ups. We further quantified the clinical agreement of the two systems using dual-reference-standard ROC analysis and performed subgroup analyses stratified by age and comorbid diabetes to identify the patients in whom this approach might add the most information. Our findings do not support presenting this tool as a more accurate classifier; they instead define what such a band can and cannot be used for in rural primary care hypertension management, and they inform the design of patient–home–hospital integrated chronic disease management systems.

## 2. METHODS

### 2.1 Study design and population

This retrospective observational cohort study was conducted in four villages in Jiasong Town, Nanhe District, Xingtai City, Hebei Province, China, between January 2023 and August 2024. A total of 200 patients with hypertension registered in the National Primary Care Chronic Disease Reporting System were consecutively enrolled, and all 200 completed the 15 scheduled visits (3,000 of 3,000 planned observations); no patient was lost to follow-up and no observation was missing.

**Inclusion criteria:** (1) confirmed diagnosis of hypertension according to national guidelines; (2) complete serial BP observation records; and (3) eligibility for inclusion under the IRB-approved waiver of written informed consent for retrospective analysis of de-identified records.

**Exclusion criteria:** (1) incomplete BP measurement records or failure to complete scheduled follow-ups; and (2) records for which reliable de-identification could not be confirmed.

Each patient completed 15 scheduled follow-up visits, yielding 3,000 valid observations. All data were de-identified by removal of personally identifiable information before analysis. Data collection was completed in August 2024, after which the retrospective protocol was submitted for ethical review; no new biological specimens were collected

and no participant was contacted. The study was approved by the Institutional Review Board (IRB) of Nanhe District People's Hospital, Xingtai, Hebei Province, China, under expedited review (approval No. NH-2024-001; approved 1 November 2024) and was conducted in accordance with the Declaration of Helsinki. Because this was a retrospective analysis of de-identified clinical records with no additional biological specimens, no interventions, and minimal risk to participants, the IRB waived the requirement for written informed consent.

Sample size was estimated a priori from pre-experimental data. Assuming an expected AUC difference of 0.14 between the two evaluation systems, with a two-sided α of 0.05 and a statistical power of 0.80, at least 178 participants were required. The final cohort of 200 patients exceeded this requirement.

### 2.2 Data collection

BP and HR were measured in the morning after the patient had rested in a seated position for at least 5 minutes. All measurements were obtained by trained operators using the same brand and batch of domestic electronic BP monitors and a standardized protocol. Three consecutive measurements were obtained at each visit, and their average was recorded as the SBP, DBP, and HR for that visit[11]. The identical protocol was applied at all 15 visits, and no additional laboratory measurement beyond BP and HR was used in the present analysis.

### 2.3 Universal hypertension standard reference

All diagnostic thresholds and treatment targets referenced in this study strictly followed the Chinese Guidelines for the Prevention and Treatment of Hypertension (2018 Revision) and the Chinese Guidelines for the Management of Hypertension in the Elderly (2023)[1,3]. The population reference intervals were defined as SBP 90–140 mmHg, DBP 60–90 mmHg, and HR 60–100 bpm.

**Important clarification:** this study did not modify any of these universal guideline thresholds. The conversion model served exclusively as an auxiliary interpretive tool, and all formal diagnostic and treatment decisions remained based on the official guideline standards.

## 2.4 Conversion model development

The population-individual reference interval conversion model was developed from the patented algorithm of Feng *et al.*[9,10]. Its architecture and core principles are described below.

### *2.4.1 Individual reference interval construction*

For each patient, the PSR-band was derived from the first 14 visits (derivation set). Visit 15 was held out as an independent validation set and was not used to estimate any band parameter. The derivation set therefore defines the band parameters only; all primary performance estimates are reported for the held-out visit 15.

- The individual mean (Mi) and individual standard deviation (SDi) were computed from the derivation set; SDi is the within-person standard deviation estimated from repeated measurements, the quantity that governs how far a single reading may depart from a patient's own level[21]. The individual band half-width (Si) was defined as 1.96 × SDi, that is, the interval range of the patient's own 95% distribution rather than a separate variance estimate.

- Population reference parameters were taken from the Chinese guideline [1]: SBP $M_p$ = 115 mmHg and $S_p$ = 25 mmHg; DBP $M_p$ = 75 mmHg and $S_p$ = 15 mmHg; and HR $M_p$ = 80 bpm and $S_p$ = 20 bpm. Here $S_p$ denotes the population reference interval half-width, which together with $M_p$ defines the population reference interval used as the rescaling anchor.

### *2.4.2 Core conversion formula*

The conversion maps a measured value from its relative position in the individual distribution to the corresponding relative position in the population reference distribution. The symbols are as follows: X, the measured value at a given visit (SBP, DBP, or HR); Mi, the individual mean of the derivation set (visits 1–14); SDi, the individual standard deviation of the derivation set; Si = 1.96 × SDi, the individual band half-width (the interval range of the patient's own 95% distribution); Mp, the population reference mean (SBP 115 mmHg, DBP 75 mmHg, HR 80 bpm); and Sp, the population reference interval half-width (SBP 25 mmHg, DBP 15 mmHg, HR 20 bpm). The conversion formula is:

$$\text{Corrected value} = \frac{\text{Measured value} - M_i}{S_i} \times S_p + M_p$$

The denominator of Z is the individual band half-width $S_i$ ($1.96 \times SD_i$; Section 2.4.1), that is, the interval range of the patient's own distribution, whereas rescaling of the corrected value uses the population reference interval half-width $S_p$. Both sides of the conversion therefore use an interval-range (half-width) metric on the same scale, so that the position of a reading within its own band is mapped consistently to the corresponding position within the population reference interval.

Corrected values—SBPc, DBPc, and HRc—were generated with this formula for individualized assessment.

*Worked example.* For a patient with SBP $M_i$ = 130 mmHg and $SD_i$ = 6 mmHg, the individual band half-width is $S_i = 1.96 \times 6 = 11.76$ mmHg. A measured value of X = 142 mmHg gives $Z = (142 - 130)/11.76 = 1.02$, and applying the population parameters ($M_p$ = 115 mmHg; $S_p$ = 25 mmHg) yields $Xc = 115 + 1.02 \times 25 = 140.5$ mmHg. The transform thus preserves the patient's within-person rank (Z-score) while rescaling it to the population reference distribution.

*Reproducibility. All symbols and formulas, together with a fully worked per-patient example and the analysis code (R 4.4), are provided in Supplementary Material S1 (Sections S1.1 and S1.6), so that every corrected value can be independently recalculated.*

*2.4.3 Methodological foundation*

This distributional mapping approach is grounded in the established statistical principles of Z-score standardization and reference interval transformation, which have been widely applied in clinical chemistry, growth assessment, and epidemiological research to normalize individual measurements against population benchmarks[9,10]. The patented algorithm used here is a tailored application of this general method to hypertension management in primary care, enabling individualized interpretation of BP values while retaining alignment with population-based guideline standards[12,13].

*Mathematical note. Because the PSR-band maps each patient's percentile rank in his or her own distribution to the corresponding percentile rank in the population distribution (population means: SBP 115 mmHg, DBP 75 mmHg, HR 80 bpm), the group mean of the corrected values is mathematically anchored to these values by construction. Three consequences follow: (a) raw sex differences disappear after correction; (b) visit-to-visit time effects on corrected values are negligible (small β; SBPc β = 0.137 mmHg per visit, P = 0.008; DBPc and HRc not significant); and (c) between-group mean differences are close to zero. These are mathematical properties of the transform and must not be cited as independent evidence of clinical validity. The clinically meaningful validation metrics are: (i) the distribution of PSR-band width across individuals; (ii) within-patient ICC and CV of the band across validation visits; (iii) the association of within-band vs. out-of-band visits with subjective tolerance and physician acceptability; and (iv) in future work, the association with ABPM circadian profile and target-organ damage. A note on the scope of the orthogonality. The within-person component u is orthogonal to the between-person level Mi only when both are computed from the same set of visits; this holds for the derivation set used here (visits 1–14) and would equally hold for any other complete set, because the deviations from a sample mean sum to zero by construction. It does not extend to a reading outside the estimating set: at the held-out visit 15 the standardised deviation is negatively correlated with Mi (r = −0.199 for SBP, −0.049 for DBP and −0.055 for HR), which is the usual regression to the mean of an out-of-sample observation and is a property of the validation point rather than of the model. The distributional statements in Section 4.4 therefore refer to the derivation set.*

### *2.4.4 Key innovation and differentiation*

Unlike traditional correction models that adjust treatment targets for specific populations, this model operates entirely within the standard threshold framework. It does not redefine what constitutes "normal" or "abnormal" by guideline definition; rather, it provides a personalized quantification of how the patient's current BP status relates to his or her own physiological steady state. All clinical decisions continue to reference universal guideline thresholds, ensuring full regulatory and clinical compliance. It thus represents an interpretive layer of personalization rather than a revision of standards.

*Distinction between guideline thresholds, population reference intervals, and the PSR-band.* These constructs serve different purposes and must not be conflated. (i) Guideline thresholds (e.g., SBP < 140 mmHg) are the fixed diagnostic and treatment criteria used for clinical decision-making. (ii) Population reference intervals (e.g., SBP 90–140 mmHg) are descriptive distributions of a reference population. (iii) The PSR-band is an individualized within-patient band describing the range within which the patient's recent BP readings tend to fall. The PSR-band is an interpretive, descriptive construct that flags deviation from the patient's own steady state; it does not by itself define "safe" or "unsafe" BP, and a reading within the band does not imply that the patient is healthy or that treatment may be de-intensified—it is merely a suggestion based on the individual's steady-state status. Establishing clinical safety or benefit requires prospective evidence against hard cardiovascular endpoints.

*2.4.5 Model properties: What the transform can and cannot add*

Within a patient the corrected value is a monotone linear function of the measured value, $Xc = Mp + (X - Mi)\cdot Sp/Si$, with the slope Sp/Si fixed for that patient. Three properties follow directly and constrain what the model can legitimately be claimed to add.

First, because the transform is monotone within each patient, it cannot alter that patient's internal ordering of readings, and any discrimination computed within a single patient is unchanged. Incremental information in a pooled analysis can therefore arise only from differences between patients—from the fact that patients are weighted by their own slope Sp/Si when pooled. This is a property of the model, not an empirical result.

Second, because $Si = 1.96\cdot SDi$ appears in the denominator, patients with a small SDi have their deviations amplified. For a patient with SDi = 2 mmHg, a change of 2 mmHg corresponds to $Z = 0.51$ and therefore to about 12.8 mmHg on the corrected SBP scale—approximately half of the population reference half-width (Sp = 25 mmHg). Such patients can be flagged as out of band on the basis of a change that is small in absolute terms. The observed distribution of SDi is reported in Section 3.2.

Third, a standard deviation estimated from 14 visits still carries sampling error: its relative standard error, approximately $1/\sqrt{(2(n-1))}$, is about 20% at $n = 14$. Because

that estimate enters the denominator directly, band width remains approximate for individual patients even when the group-level distribution is well estimated. The structural consequence of this weighting is derived in the companion methodological paper[23].

The workflow of the population-individual reference interval conversion model is illustrated in Figure 1.

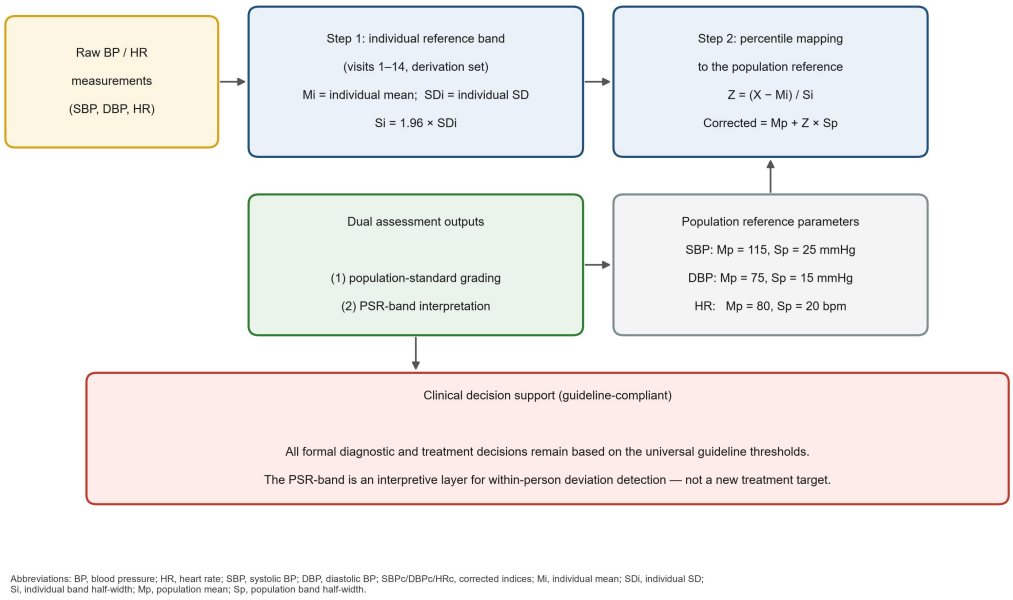


**Figure 1.** Workflow of the personal steady-state reference band (PSR-band) conversion model. Raw BP and HR measurements are first used to construct an individual reference band (Mi ± 1.96·SDi) from visits 1–14; subsequent measurements are Z-mapped to the population reference (Mp ± Sp) to yield the corrected indices (SBPc/DBPc/HRc). The PSR-band serves as an interpretive layer, and all formal diagnostic and treatment decisions remain based on universal guideline thresholds.

### 2.5 Outcome measures

**Primary outcome:** grading consistency between population-based standard assessment and individualized corrected assessment, evaluated with Cohen's kappa and Cramér's V.

**Secondary outcomes:**

1. Magnitude of the reduction in baseline differences by sex after correction
2. Temporal stability of measured *vs.* corrected values across 15 serial follow-ups
3. Agreement with comprehensive clinical judgment, evaluated by ROC-AUC
4. Subgroup differences in grading inconsistency rates stratified by age and comorbid diabetes

**2.6 Statistical analysis**

All statistical analyses were performed using R 4.4 (lme4, geepack) and SPSS 29.0 (IBM Corp., Armonk, NY, USA). Because the 15 visits were nested within 200 patients, continuous outcomes (SBP, DBP, HR, and corrected indices) were analyzed with linear mixed-effects models (LMM) including a random intercept for patient: $y_{ij} = \beta_0 + \beta_1 \cdot time + \beta_2 \cdot sex + b_i + \varepsilon_{ij}$, estimated by restricted maximum likelihood (REML). Treatment regimen was not recorded in a form suitable for inclusion as a covariate, and a single-cohort design has no group-allocation variable, so the model was restricted to time and sex as fixed effects. Binary outcomes (e.g., whether a visit fell within the PSR-band) were analyzed with generalized estimating equations (GEE) using an exchangeable correlation structure and robust standard errors. Cohen's kappa and Cramér's V were used to assess between-system grading consistency. Agreement bands follow Landis and Koch[18] (< 0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, > 0.80 almost perfect); a pre-specified three-band scheme (≥ 0.75 good; 0.40–0.74 moderate; < 0.40 poor) was also applied so that interpretation would not rest on a single taxonomy. All observed values fall below 0.40 under both schemes. A simulation-based sensitivity analysis of how the number of derivation visits affects band width is reported in Supplementary Material S1 (Section S1.5). Two-sided $P < 0.05$ was considered statistically significant. No formal multiplicity correction was applied: the study is exploratory and hypothesis-generating, the outcome families were not pre-specified in a hierarchical testing plan, and the ROC comparisons rest on a single cross-sectional assessment. *P* values are therefore reported descriptively rather than as confirmatory evidence.

Dual-reference-standard ROC analysis was used to evaluate the agreement of the two evaluation systems with clinical judgment. ROC curves were constructed with patient subjective tolerance and physician comprehensive clinical judgment as reference standards; the AUC and its 95% CI were calculated for each index. Because

discrimination rather than threshold selection was the question of interest, operating points are not reported. Because the indices were compared within the same patients, the paired DeLong test for correlated ROC curves was used and the AUC difference is reported with its 95% confidence interval and two-sided *P* value [Table 2]. The anchor definition is stated explicitly because it determines the direction of every AUC: the positive class was a rating of 5 on the five-point physician and patient scales, which denotes a reading judged acceptable (physician) or tolerated (patient); higher ratings were the negative class. All 200 readings carried a definitive rating, so no reclassification was required. All ROC analyses use the visit-15 cross-section and are exploratory.

### 2.7 Clinical acceptability anchors for ROC analysis (Not hard endpoints)

A blinded, independent evaluation design was used for ROC validation. The two anchors—physician clinical acceptability and patient subjective tolerance—are feasibility measures of cross-sectional agreement; they are not hard clinical endpoints and do not by themselves establish that the PSR-band defines a safe or optimal BP level. Accordingly, the ROC analyses in this study are exploratory and hypothesis-generating, and they should not be cited as evidence that the PSR-band identifies safe or optimal BP levels.

Physician comprehensive clinical judgment: three board-certified general practitioners with at least 5 years of experience in primary care chronic disease management independently rated each patient's BP status on a five-point scale coded 5 to 9, where 5 denotes a reading judged fully acceptable and 9 a reading judged unacceptable; ratings took account of the baseline BP trajectory, comorbidity, treatment history, symptoms and organ perfusion. Patient subjective tolerance was rated by the patient on the same five-point scale on the basis of dizziness, headache and fatigue. The positive (reference) class for both anchors was a score of 5, giving 158 positive and 42 negative patient ratings and 166 positive and 34 negative physician ratings; every one of the 200 readings had a definitive rating.

**Patient subjective tolerance:** patients rated their subjective tolerance of their current BP status on the same 5-point Likert scale, based on the presence of symptoms such as dizziness, headache, and fatigue.

During evaluation, both physicians and patients were blinded to the results of the individualized correction model and were provided only with raw BP values and basic clinical information to preserve objectivity. All 200 patients received a definitive rating; the numbers of positive and negative ratings under each anchor are reported in Section 3.6.

## 3. RESULTS

### 3.1 Baseline characteristics

A total of 200 patients with hypertension were included in the final analysis (79 male and 121 female), with a mean age of 66.14 ± 8.88 years. Thirty-nine patients had comorbid diabetes mellitus, and the remaining 161 had hypertension without diabetes. At the held-out visit 15, mean SBP was 133.28 ± 9.07 mmHg, mean DBP 80.12 ± 7.98 mmHg and mean HR 71.08 ± 7.33 bpm. Across the cohort, the patient-specific means (Mi) derived from visits 1–14 were 134.03 ± 6.59, 80.24 ± 4.75 and 70.86 ± 4.39 (mmHg, mmHg and bpm), and the corresponding standard deviations (SDi) were 7.94 ± 3.99, 5.76 ± 1.87 and 4.44 ± 1.76. Over the whole study period, 3000 valid observations were collected (1185 from male and 1815 from female patients).

### 3.2 Overall comparison of measured and corrected indices

On the held-out visit-15 cross-section ($n$ = 200), mean measured SBP was 133.28 ± 9.07 mmHg and mean corrected SBP (SBPc) 114.93 ± 16.41 mmHg; mean measured DBP was 80.12 ± 7.98 mmHg and mean DBPc 74.32 ± 10.23 mmHg; and mean measured HR was 71.08 ± 7.33 bpm and mean HRc 80.95 ± 13.33 bpm. The individual band half-widths (Si = 1.96 × SDi) were 15.6 ± 7.4 mmHg for SBP (median 13.3; interquartile range 9.8–19.8; range 3.6–41.5), 11.3 ± 3.1 mmHg for DBP (10.9; 8.9–13.1; 3.6–26.7) and 8.7 ± 3.4 bpm for HR (8.2; 6.2–10.8; 3.3–21.2). Because Si enters the conversion as a denominator, the transform acts far more strongly on patients for whom the band is narrow, which is the source of the between-patient re-weighting described in Section 2.4.5.

### 3.3 Sex differences and temporal stability

**Baseline differences by sex:** raw measured values showed the expected physiological differences (men had higher DBP and women had higher HR; all $P < 0.001$ by LMM on

raw values). After individualized correction, no statistically significant sex differences remained in SBPc, DBPc, or HRc (all $P > 0.96$). Because the transform maps every patient onto the same population reference distribution, disappearance of the raw sex differences is expected by construction rather than being an independent clinical finding (Mathematical note, Section 2.4.3).

Temporal stability across follow-ups: LMM on raw values confirmed significant visit-to-visit variation in measured SBP and DBP. For corrected values, the visit effect on SBPc was small ($\beta = 0.137$ mmHg per visit, $P = 0.008$) and not significant for DBPc ($P = 0.785$) or HRc ($P = 0.076$). These time effects on corrected values are negligible relative to the residual SD (approximately 12 mmHg for SBPc) and are not interpreted as "noise filtering"; the clinically meaningful stability metrics are the band-width distribution reported in Section 3.2 and the within-patient reproducibility summarised in Supplementary Material S1.

Between the first and last visits (visits 1 and 15), DBPc changed from its derivation-set value to 74.32 ± 10.23 mmHg at the held-out visit. This within-patient trend is hypothesis-generating and requires longer follow-up; the band is therefore presented as a description of the recent steady state, not as a fixed individual target.

### 3.4 Grading consistency between the two evaluation systems

When BP status was graded by population fixed thresholds *vs.* individualized correction, the two systems showed poor agreement by the predefined criteria. On the held-out visit 15, Cohen's kappa was 0.242 at 140 mmHg and 0.112 at 130 mmHg for SBP, and 0.307 at 90 mmHg and 0.266 at 80 mmHg for DBP [Table 1]. All four values lie below 0.40 and are therefore "poor" under the prespecified scheme and "fair" under the Landis–Koch scheme[18]. The gap between the two systolic thresholds is itself informative: the lower the threshold, and the closer it lies to what most members of an already-treated cohort can reach, the poorer the agreement.

Among patients classified as normal by the population standard, 91.4% (SBP) and 91.8% (DBP) of readings also lay within the individual band, so the two systems agree on most normal readings. Among the 49 patients classified as SBP-abnormal by population criteria, 40 (81.6%) fell within their own band and were therefore not

flagged by the individualized assessment; the corresponding figure for the 29 DBP-abnormal patients was 22 (75.9%). This population-abnormal/within-band state is the numerically dominant form of discordance and is the one that must not be read as grounds for de-intensification.

In the integrated fusion framework, the joint reading of the two systems produced concordant-abnormal and concordant-normal states that increase diagnostic certainty, together with two discordant states that require explicit caution. The discordant population-abnormal/within-band state was the more common of the two: 81.6% of population-abnormal SBP readings and 75.9% of population-abnormal DBP readings lay within the patient's own band. The reverse state—population-normal but outside the band—was less common (6.5% and 7.0% of readings) but carries a different message: the reading has moved away from the patient's recent steady state even though it meets the guideline threshold.

**Table 1. Grading consistency between the population standard and individualized correction at two thresholds each for SBP and DBP (visit 15, *n* = 200).**

| Indicator | Threshold | Concordant cases | Overall agreement | Cohen's kappa | Cramér's V | *P* value |
|---|---|---|---|---|---|---|
| SBP | 140 mmHg | 159 | 79.5 | 0.242 | 0.349 | < 0.001 |
| SBP | 130 mmHg | 81 | 40.5 | 0.112 | 0.244 | < 0.001 |
| DBP | 90 mmHg | 175 | 87.5 | 0.307 | 0.362 | < 0.001 |
| DBP | 80 mmHg | 120 | 60.0 | 0.266 | 0.347 | < 0.001 |

Note. BP = blood pressure; SBP = systolic blood pressure; DBP = diastolic blood pressure. Both systems are classified against the same upper threshold: a corrected reading is counted as abnormal when the corrected value equals or exceeds that threshold, which is not the same as falling outside the population reference interval (a distinction that matters at the 130 mmHg and 80 mmHg thresholds). *P* values are from Pearson chi-square tests. Kappa and Cramér's V are computed directly from the four cells.

### 3.5 Subgroup analysis

Stratified analyses by age and by comorbid diabetes (39 patients) are shown in Figure 2. Across the whole cohort the between-system inconsistency rate at the held-out visit was 26.5% (53/200) for systolic pressure and 18.0% (36/200) for diastolic pressure.

- Age stratification: the systolic rate was identical in patients aged ≥ 65 and < 65 years (26.5% *vs.* 26.5%), whereas the diastolic rate was lower in the older group (12.9% *vs.* 27.9%). In this cohort, therefore, age did not identify a subgroup with systematically greater disagreement about systolic pressure, and moved in the opposite direction for diastolic pressure.

- Diabetes stratification: patients with diabetes showed a substantially higher inconsistency rate than those without (SBP 41.0% *vs.* 23.0%; DBP 25.6% *vs.* 16.1%). The diabetes subgroup ($n$ = 39) is adequately sized for descriptive comparison, but formal inferential subgroup analysis remains underpowered.
-

Diabetes, rather than age, therefore marks the subgroup in which the two standards diverge most; whether individualized interpretation adds information there remains to be tested. These subgroup differences are visualized in Figure 2.

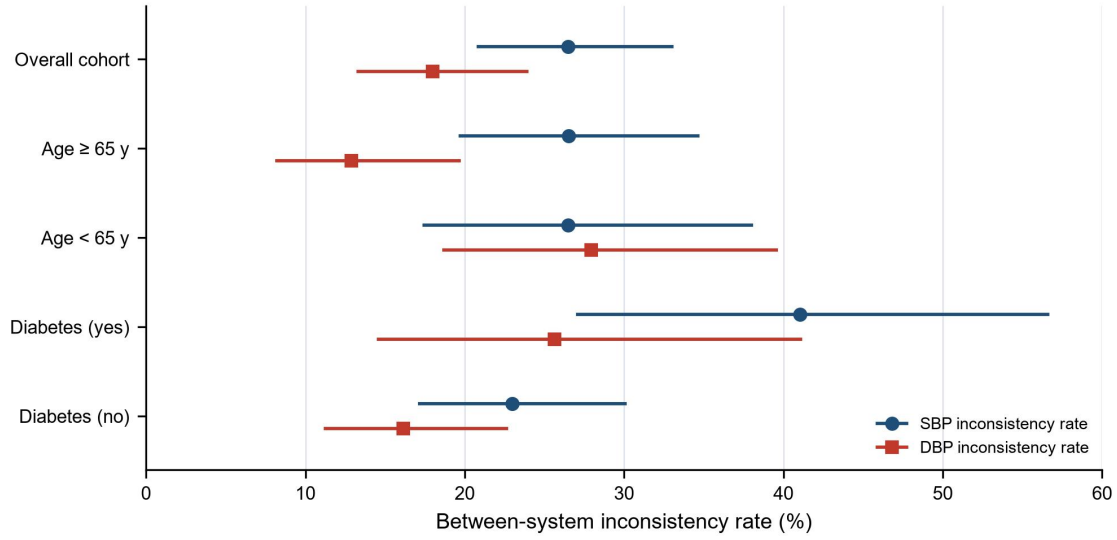


**Figure 2.** Between-system inconsistency rate at the 140 mmHg (systolic) and 90 mmHg (diastolic) thresholds, by age and diabetes status. Points are rates and bars Wilson 95% confidence intervals.

### 3.6 Model performance: Agreement with clinical judgment

*3.6.1 Patient subjective tolerance as the reference standard*

On the held-out visit 15 (200 patients; 158 rated the reading as tolerated and 42 as not tolerated), ROC analysis with patient tolerance as the reference standard gave AUCs of 0.655 ($P = 0.003$) for SBP, 0.829 ($P \leq 0.001$) for DBP, 0.492 ($P = 0.875$) for HR, 0.661 ($P = 0.002$) for SBPc, 0.812 ($P \leq 0.001$) for DBPc and 0.536 ($P = 0.459$) for HRc. Diastolic pressure discriminated strongly, systolic pressure moderately, and heart rate not at all.

Paired DeLong testing showed no advantage for the corrected indices: ΔAUC was +0.006 for SBP (95% CI -0.064 to 0.075, $P = 0.871$), -0.017 for DBP (-0.076 to 0.041, $P = 0.556$) and +0.044 for HR (-0.033 to 0.122, $P = 0.264$). The absence of an independent objective anchor (for example the ambulatory BP circadian profile or target-organ damage) remains a key limitation.

*3.6.2 Physician comprehensive clinical judgment as the reference standard*

With physician judgment as the reference standard on the same held-out cross-section (166 readings judged acceptable and 34 not acceptable), AUCs were 0.901 ($P \leq 0.001$) for SBP, 0.658 ($P = 0.004$) for DBP, 0.479 ($P = 0.663$) for HR, 0.843 ($P \leq 0.001$) for SBPc, 0.648 ($P = 0.005$) for DBPc and 0.510 ($P = 0.853$) for HRc. Systolic pressure discriminated almost perfectly in its measured form, and correction reduced rather than preserved this.

Paired comparisons confirmed that the corrected indices were not better and, for systolic pressure, were worse: ΔAUC was -0.058 for SBP (95% CI -0.102 to -0.014, $P = 0.010$), -0.010 for DBP (-0.090 to 0.070, $P = 0.809$) and +0.031 for HR (-0.060 to 0.123, $P = 0.502$). The pattern across both anchors is therefore one of no gain, and in one channel a measurable loss.

The AUC comparison in Table 2 and the corresponding ROC curves in Figure 3 are presented on this basis.

**Table 2. Measured and corrected indices against the two clinical reference standards (visit 15, *n* = 200): areas under the curve, paired differences and DeLong tests.**

| Reference standard | Index | Measured AUC (95% CI) | Corrected AUC (95% CI) | ΔAUC, corrected − measured (95% CI) | DeLong *P* |
|---|---|---|---|---|---|
| Patient tolerated (158 / 42) | SBP | 0.655 (0.553–0.757) | 0.661 (0.559–0.762) | +0.006 (-0.064 to 0.075) | 0.871 |
| | DBP | 0.829 (0.755–0.904) | 0.812 (0.738–0.885) | -0.017 (-0.076 to 0.041) | 0.556 |
| | HR | 0.492 (0.394–0.590) | 0.536 (0.440–0.633) | +0.044 (-0.033 to 0.122) | 0.264 |
| Physician judgment (166 / 34) | SBP | 0.901 (0.829–0.973) | 0.843 (0.761–0.925) | -0.058 (-0.102 to -0.014) | 0.010 |
| | DBP | 0.658 (0.551–0.765) | 0.648 (0.545–0.752) | -0.010 (-0.090 to 0.070) | 0.809 |
| | HR | 0.479 (0.383–0.574) | 0.510 (0.402–0.619) | +0.031 (-0.060 to 0.123) | 0.502 |

*Note. AUC = area under the curve; CI = confidence interval; DeLong *P* = paired test for correlated ROC curves. Values are continuous ROC AUCs on the visit-15 cross-section (*n* = 200). The positive (reference) class is a rating of 5 on the five-point physician and patient scales (acceptable or tolerated); higher ratings form the negative class. ΔAUC = corrected − measured; its 95% CI and *P* value come from the paired DeLong test. For reference, the fully in-sample (all-fifteen-visit) estimate for systolic pressure against physician judgment is 0.840 (*P* = 0.007); every other comparison is unchanged in direction (Section 3.6.3).

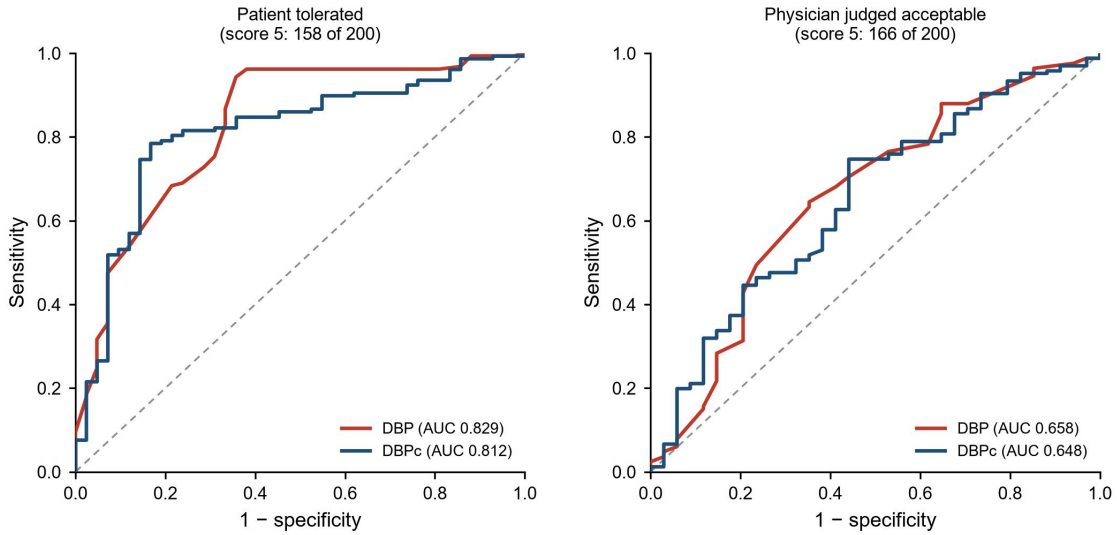


**Figure 3.** Receiver operating characteristic curves for measured and corrected diastolic pressure on the visit-15 cross-section. The positive class is a rating of 5: (A) readings the patient tolerated (158 of 200) and (B) readings the physician judged acceptable (166 of 200). Red, measured DBP; blue, corrected DBPc. Areas under the curve are given in the legends; the dashed diagonal is the line of no discrimination.

*3.6.3 Interpretation of results*

Three observations deserve emphasis. First, the measured indices are informative: diastolic pressure tracked patient tolerance (AUC 0.829) and systolic pressure tracked physician judgment (AUC 0.901). Whatever the PSR-band adds, it is not needed to establish that the raw readings carry clinical signal.

Second, correction did not improve either index, and for systolic pressure against physician judgment it significantly reduced discrimination (0.901 to 0.843, $P = 0.010$). The model's mathematics accounts for this. Because each patient is rescaled by his or her own slope Sp/Si, patients with small within-person variability are spread across the corrected scale while those with large variability are compressed toward the population mean; the corrected ordering therefore no longer follows the absolute values on which the physicians were judging, and discrimination falls.

Third, heart rate behaved differently from pressure: the corrected index was numerically better under both anchors (0.536 *vs.* 0.492 and 0.510 *vs.* 0.479) without reaching significance, which suggests the transform may add information where the measured

index itself carries none.

Sensitivity analysis: the fully in-sample case. When the band is instead estimated from all fifteen visits and the same fifteen readings are evaluated, the qualitative pattern is unchanged. Systolic discrimination against physician judgment falls from 0.901 to 0.840 (paired DeLong $P = 0.007$) rather than to 0.843 ($P = 0.010$), and the four kappa values are 0.146, 0.091, 0.235 and 0.268 rather than 0.242, 0.112, 0.307 and 0.266; both sets remain below 0.40. The in-sample design is the one to which Proposition 1 of the companion paper applies exactly, and its agreement with the held-out design means that the conclusion does not rest on the choice of validation scheme. We report the held-out analysis as primary because the band is intended for use on readings not yet observed.

**3.7 Relationship between the individual band and the population reference interval**

No band coincided with the population interval (0/200 for each of the three indices; exact coincidence would require Mi = Mp and SDi = Sp/1.96 simultaneously) and no band was fully disjoint from it (0/200). Every band therefore either lay wholly inside the population interval—or, in a minority of patients, contained it—or crossed its boundary [Table 3]. For SBP, 27.0% of bands were nested and 73.0% partially overlapped; the corresponding figures were 44.0% and 56.0% for DBP and 64.0% and 36.0% for HR. The ordering across indices follows the size of within-person variability: HR has the narrowest bands and hence the largest proportion of nested bands.

At the level of individual readings the two systems assigned different categories to 26.5% of systolic, 18.0% of diastolic and 11.5% of heart-rate readings [Table 4]. Most disagreement was a reading above the guideline threshold that nevertheless lay within the patient's own band (20.0%, 11.0% and 0.0% respectively); the reverse discordance—within threshold but outside the band—was less common for pressure (6.5% and 7.0%) but accounted for all of the discordance seen for heart rate (11.5%).

The direction of discordance differs between the two channels, and so does its mechanism. For systolic and diastolic pressure the dominant form is a reading above the threshold that nevertheless lies within the band, and it occurs preferentially in patients with a wide band: the median band half-width was 17.3 mmHg in discordant and 12.2 mmHg in concordant patients (21.1 mmHg in the 40 patients with an above-threshold

but within-band reading). Heart rate is the opposite case: all of its discordance consists of readings within the threshold but outside the band, and it arises in patients with a narrow band who deviated substantially (median half-width 7.0 bpm versus 8.4 bpm in concordant patients; median |u| 1.23 versus 0.38). In other words, the pressure channels err by making the band too wide and the heart-rate channel by making it too narrow.

**Table 3. Relationship between the population reference interval and the individual band (*n* = 200 patients).**

| Index | Coincident, *n* (%) | Nested, *n* (%) | Partial overlap, *n* (%) | Fully disjoint, *n* (%) |
|---|---|---|---|---|
| SBP | 0 (0) | 54 (27.0) | 146 (73.0) | 0 (0) |
| DBP | 0 (0) | 88 (44.0) | 112 (56.0) | 0 (0) |
| HR | 0 (0) | 128 (64.0) | 72 (36.0) | 0 (0) |

Note. Nested = the band lies wholly inside the population interval (or, in 6 DBP and 1 HR patient, the population interval lies wholly inside the band). Coincident was assessed with a ±1 mmHg tolerance; exact coincidence would require Mi = Mp and SDi = Sp/1.96 at the same time and is therefore not expected to occur.

**Table 4. Classification of the 200 held-out readings by the population standard and the individual band.**

| Index | Both normal | Above threshold, within band | Within threshold, outside band | Both abnormal | Discordant, *n* (%) |
|---|---|---|---|---|---|
| SBP | 138 (69.0) | 40 (20.0) | 13 (6.5) | 9 (4.5) | 53 (26.5) |
| DBP | 157 (78.5) | 22 (11.0) | 14 (7.0) | 7 (3.5) | 36 (18.0) |
| HR | 176 (88.0) | 0 (0.0) | 23 (11.5) | 1 (0.5) | 23 (11.5) |

Note. Values are *n* (%) of the 200 readings at the held-out visit 15. Discordant = column 3 + column 4: 26.5% for SBP, 18.0% for DBP and 11.5% for HR.

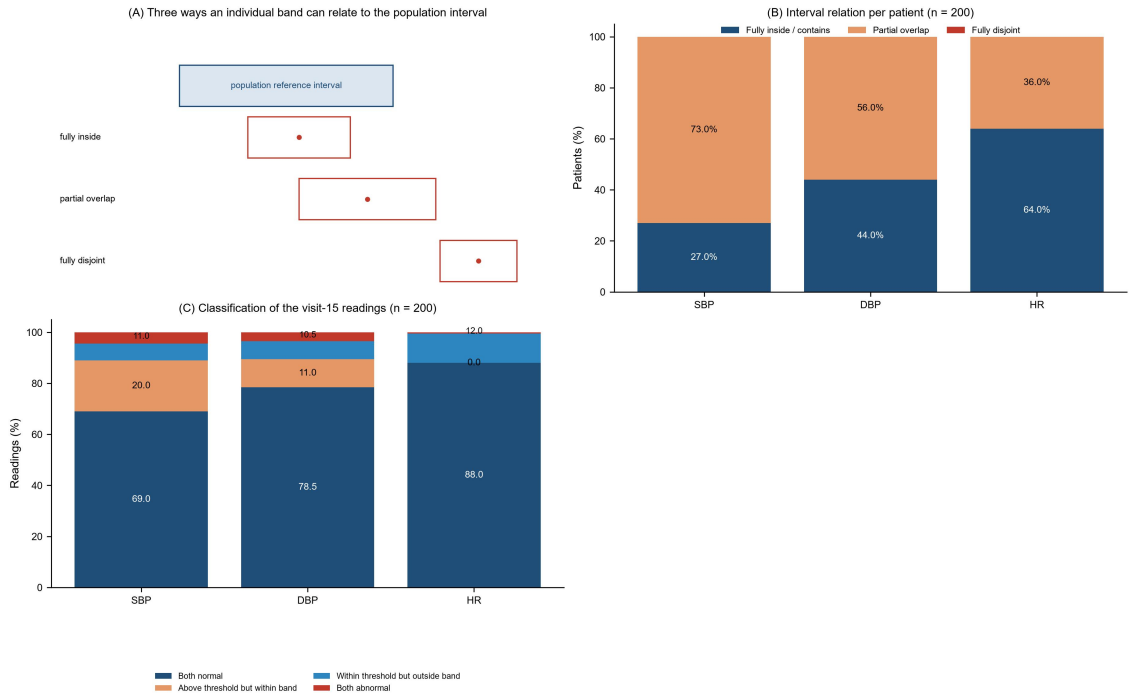


**Figure 4.** Concordance between the population reference interval and the individual band. (A) The three possible interval relations; (B) the relation observed in each of the 200 patients; (C) classification of the 200 held-out readings by the two systems.

## 4. DISCUSSION

### 4.1 Principal findings

This study derived a personal steady-state reference band (PSR-band) and evaluated it in a real-world rural primary care cohort of 200 patients with hypertension. The principal findings are as follows. (1) The band operates entirely within universal guideline thresholds without modifying any diagnostic or treatment standard, and is therefore an interpretive layer rather than a new target. (2) The anchoring of corrected indices to the population means, the disappearance of raw sex differences and the null visit-to-visit time effects are mathematical properties of the mapping, not independent clinical findings (Section 2.4.3). (3) Grading agreement between the population standard and the band is poor (Cohen's kappa 0.112–0.307). (4) On the held-out visit 15 the measured indices discriminated strongly (DBP 0.829 against patient tolerance; SBP 0.901 against physician judgment), whereas the corrected indices did not improve on them—ΔAUC −0.017 for DBP ($P$ = 0.556) and +0.006 for SBP ($P$ = 0.871) against patient tolerance—and for SBP against physician judgment correction was significantly worse (0.843 *vs.* 0.901, $P$ = 0.010). (5) Between-system disagreement was higher in patients with diabetes (SBP 41.0% *vs.* 23.0%; DBP 25.6% *vs.* 16.1%) but showed no

systolic age gradient, so comorbidity rather than age identifies the subgroup in which the two standards diverge. (6) No individual band coincided with the population reference interval (0/200) and none was fully disjoint from it (0/200); bands were either nested (27.0% for SBP) or partially overlapping (73.0%).

The two evaluation systems should be viewed as complementary rather than competing: the population standard provides an evidence-based long-term compliance anchor, whereas the individualized correction reflects current steady-state appropriateness. Together they form a dual-assessment paradigm that balances guideline adherence with personalized care.

### 4.2 Integrated fusion framework: Concordance, discordance, and the need for caution

The two evaluation systems are best understood through their fusion into an integrated framework. For each visit, the population standard and the individualized PSR-band assessment jointly generate four concordance states: (i) concordant-abnormal (both systems flag the reading as abnormal or outside the band); (ii) concordant-normal (population-normal and within-band); (iii) discordant population-abnormal/within-band (the population standard flags the reading as abnormal while it falls within the patient's own steady-state band); and (iv) discordant population-normal/outside-band (the population standard considers the reading normal while it lies outside the patient's own band).

Discordance—in either direction—is the clinically decisive signal and must be handled with explicit caution. For state (iii), a population-classified abnormal reading that nonetheless falls within the patient's own band does not rule out a clinically meaningful elevation and must not be used as grounds for treatment de-intensification. Rather, it calls for a prudent, individualized review: reconfirm the reading, evaluate symptoms, comorbidities, tolerance, and dynamic trends, and intensify monitoring rather than mechanically adjusting therapy. For state (iv), a population-normal reading that lies outside the patient's own band signals a possible departure from that patient's recent steady state and warrants trend review and reconfirmation; it should not be dismissed merely because it meets the population threshold.

In every discordant scenario, clinical decisions should return to the universal guideline framework and be integrated with comprehensive clinical judgment (symptoms, comorbidity, tolerance, ambulatory BP profile, and target-organ status); a single statistically derived band should never, by itself, justify either de-intensification or escalation. Concordant states increase diagnostic certainty, whereas discordant states flag the need for individualized, cautious management.

The framework therefore organizes the two descriptions rather than merging them: the population standard supplies the long-term anchor, and the PSR-band supplies the current steady-state reading. The null discriminative result reported above does not undermine this organizing role, because the framework claims descriptive, not predictive, value.

### 4.3 Comparison with the existing literature

Our dual-reference ROC analysis, with an explicit anchor definition and a paired DeLong test, found no advantage for the corrected indices over the measured values and a significant disadvantage for systolic pressure under physician judgment (0.901 to 0.843, $P = 0.010$). This is worth stating plainly, because the appeal of individualized indices rests on an implicit assumption that they discriminate better; for pressure in this cohort they do not, and their value, if any, lies in describing within-person position rather than in ranking patients more accurately.

### 4.4 Mechanistic insights

The mechanistic basis of the model is distributional mapping: a patient's percentile rank in his or her own derivation distribution is mapped to the corresponding percentile rank of the population reference distribution. This formalizes what experienced clinicians do implicitly—considering baseline status, treatment history, and patient tolerance rather than applying strictly threshold-based criteria. The two anchors attached to different channels: diastolic pressure tracked patient tolerance (0.829) whereas systolic pressure tracked physician judgment (0.901). That dissociation is clinically plausible—patients report symptoms such as dizziness and fatigue that are sensitive to diastolic and pulsatile load, while clinicians reason from systolic values and comorbidity—and it argues for reporting both anchors rather than selecting one.

The two-component decomposition makes the same point geometrically. Measured readings form a cap-shaped joint density over the within-person position u and the between-person level Mi, with the two axes orthogonal. Correction rescales every patient to the same mean (Mp) and the same standard deviation (Sp/1.96), so the between-person axis collapses: between-patient SD falls from 6.45 to 0.006 mmHg for SBP, 4.76 to 0.004 mmHg for DBP and 4.40 to 0.008 bpm for HR, and the intraclass correlation of the corrected values is 0.000 for all three indices[19,20]. The cap becomes a wall. Because discriminative power between patients depends on exactly the axis that correction removes, the fall in AUC for systolic pressure is structural rather than sample-specific. The same geometry explains why the two intervals almost never coincide and never fully separate: the band retains a level of its own, but it is never identical to the population interval, and it is never far enough away to lose contact with it. These statements are made for the derivation set, where the orthogonality of u and Mi is exact; the held-out visit 15 is a single out-of-sample evaluation and is analysed separately in Section 3.6. The proof of this statement, and the simulation that isolates it from any one cohort, are given in the companion methodological paper[23].

### 4.5 Clinical implications

ROC analysis on the held-out visit showed that the PSR-band agreed with the clinical anchors to about the same degree as the measured values, and no better. Approximately 83%–90% of population-classified abnormal readings fell within the patient's own steady-state band, suggesting that a substantial proportion of "abnormal" readings may represent within-person variation rather than a true clinically meaningful deviation. The PSR-band should be understood as a dynamic, step-by-step monitor of current status relative to the patient's own recent baseline—analogous to walking up or down stairs, where each step is evaluated individually rather than against a single fixed target—and it is continuously updated as more data accumulate. Individualized assessment may help clinicians identify patients whose current BP is within their own recent steady state and, pending prospective evidence, inform more gradual titration. The model can be integrated into home–hospital integrated chronic disease management systems, in which readings from validated home monitors, obtained according to standardised home-monitoring protocols[17], are entered into the system and the system returns both population-standard and PSR-band interpretations for clinician review. These clinical implications are hypothesis-generating and must be confirmed in prospective studies

with hard cardiovascular outcomes before any clinical use.

### 4.6 Limitations

First, this was a single-center retrospective study in a rural, elderly hypertensive population in northern China, and generalizability to other settings is limited. The band was derived from visits 1–14 and validated on the single held-out visit 15, which prevents the circularity of estimating and testing on the same visit, although the derivation set still contributes fourteen of the fifteen available readings.

Second, the anchors are single ratings on a five-point scale collapsed to acceptable versus not acceptable; collapsing loses information, and no independent adjudication of the ratings was available, so the comparison rests on the clinicians' and patients' global impressions rather than on a structured instrument.

Third, the PSR-band was derived from 14 baseline visits and validated on a single held-out visit. With 14 observations per patient the relative standard error of SDi is about 20%, and because SDi enters the transform as a denominator this imprecision propagates into band width and therefore into the probability of being classified as outside the band. A lower bound on SDi or empirical-Bayes shrinkage would be a reasonable refinement; neither is applied here, so individual-level band widths should be read as approximate.

Finally, the a priori sample size calculation assumed a between-system AUC difference of 0.14, whereas the paired differences ranged from −0.061 to +0.047, so the cohort has limited power for small effects; the null paired results should therefore be read as excluding a large advantage rather than as proving exact equivalence.

## 5. CONCLUSION

In conclusion, the PSR-band provides a guideline-compliant interpretive layer for within-person BP monitoring in rural primary care. It did not discriminate better than the measured readings from which it is derived, and for systolic pressure judged by clinicians it discriminated significantly worse; the present feasibility data therefore do not establish that the band defines a safe or optimal BP level, nor that de-intensification guided by the band is safe. A prospective randomised trial with hard cardiovascular

endpoints is required before clinical implementation beyond an interpretive aid.

## DECLARATIONS

### Acknowledgments

The authors thank the clinical staff of Nanhe District People's Hospital for their support in data collection and clinical evaluation, and the primary care physicians and community health workers of the village clinics in Jiasong Town for their assistance with data collection and follow-up. We are grateful to all participants in this cohort.

### Authors' contributions

Study conception and design, patented algorithm model design, overall supervision, and manuscript review: Feng L;

Clinical data collection and curation, statistical interpretation, and manuscript drafting: Zhang X, Yao J;

Clinical data collection and verification: Song J, Feng Y, Yang X;

Subgroup statistical analysis: Feng Y, Yang X, Xu Y, Yan J;

Statistical interpretation and manuscript revision: Liu W.

### Availability of data and materials

The de-identified dataset and the analysis code (R 4.4) that support the findings of this study are available from the corresponding author on reasonable request.

### AI and AI-assisted tools statement

All figures were drawn by the authors directly from de-identified individual-level data using Python 3.12 and matplotlib 3.8; no generative-AI tool was used to create or render any figure. Generative AI was not used for data analysis, interpretation, or the drawing of scientific conclusions.

### Financial support and sponsorship

This work was supported by the Medical Science Research Project of the Hebei Provincial Health Commission (No. 20240963). The funder had no role in study design, data collection, statistical analysis, manuscript preparation, or the decision to submit for publication.

## Conflicts of interest

Lianyuan Feng is the inventor and holder of the Chinese granted patents CN106326649B (granted 12 July 2024) and CN106339433B (granted 24 August 2021) underlying the reference interval conversion model. The model is presented here solely as a research method; it has not been validated, licensed or approved as a diagnostic or therapeutic tool and has no approved clinical indication. The authors declare no commercial licensing income or revenue, and no institution or funder with a commercial interest in the model was involved in this study. The remaining authors declare no conflicts of interest.

## Ethical approval and consent to participate

The study was approved by the Institutional Review Board of Nanhe District People's Hospital, Xingtai, Hebei Province, China, under expedited review (approval No. NH-2024-001; approved 1 November 2024) and was conducted in accordance with the Declaration of Helsinki. The approval document and its English translation accompany this submission. Data collection was completed in August 2024, after which the retrospective protocol was submitted for ethical review; no new biological specimens were collected and no participant was contacted. Because the analysis used de-identified clinical records with no additional specimens, no interventions and minimal risk, the requirement for written informed consent was waived by the IRB under the Measures for Ethical Review of Biomedical Research Involving Human Subjects of China.

## Consent for publication

Not applicable. The study reports only aggregate, de-identified data; individual details, images or recordings of any participant are not published, and no information that could identify an individual is included in the manuscript or in any supplementary file.

## Copyright

© The Author(s) 2026.

## REFERENCES

1. Chinese Hypertension Prevention and Treatment Guidelines Revision Committee, Hypertension Alliance (China), Chinese Society of Cardiology, et al. Chinese guidelines for the prevention and treatment of hypertension (2018 revision). *Chin J Cardiovasc*

*Med*. 2019;24:24-56.

2. Wang Z, Chen Z, Zhang L, Wang X, Hao G, Zhang Z, et al. Status of hypertension in China: results from the China Hypertension Survey, 2012-2015. *Circulation*. 2018;137:2344-56.[DOI: 10.1161/CIRCULATIONAHA.117.032380]

3. Chinese Geriatric Society Hypertension Branch, Beijing Hypertension Association, National Clinical Research Center for Geriatric Diseases, et al. Chinese guidelines for the management of hypertension in the elderly (2023). *Chin J Hypertens*. 2023;31:508-38.

4. Unger T, Borghi C, Charchar F, Khan NA, Poulter NR, Prabhakaran D, et al. 2020 International Society of Hypertension global hypertension practice guidelines. *Hypertension*. 2020;75:1334-57.[DOI: 10.1161/HYPERTENSIONAHA.120.15026]

5. Zhang YQ. Individualized treatment strategies and new advances in hypertension in the elderly. *Chin J Cardiovasc Med*. 2025;30:109-14.

6. Chinese Hypertension League, Committee of the Expert Recommendations for High-Quality Blood Pressure Management in Hypertensive Patients. Expert recommendations for high-quality blood pressure management in hypertensive patients. *Chin J Hypertens*. 2024;32:104-11.[DOI: 10.16439/j.issn.1673-7245.2024.02.002]

7. Chinese Medical Association, Chinese Society of General Practice, et al. Guideline for grassroots diagnosis and treatment of hypertension (2019). *Chin J Gen Pract*. 2019;18:301-13.[DOI: 10.3760/cma.j.issn.1671-7368.2019.04.002]

8. National Health Commission of the People's Republic of China. Guidelines for the prevention and treatment of cerebrovascular diseases (2024 edition). Beijing: National Health Commission; 2024.

9. Feng LY, Feng S. Real-world individual medical auxiliary information interaction conversion platform system and application based on evidence-based medicine. China Patent CN106326649B. 2017-01-11 (granted 2024-07-12).

10. Feng LY, Feng S. Method and device for interactive comparison platform based on population data and individual data. China Patent CN106339433B. 2017-01-18 (granted 2021-08-24).

11. Chinese Hypertension League, Committee of the Guidelines for Ambulatory Blood Pressure Monitoring. 2020 Chinese guidelines for ambulatory blood pressure monitoring. *Chin J Front Med Sci (Electron Ed)*. 2021;13:34-51.

12. Horn PS, Pesce AJ. Reference intervals: an update. *Clin Chim Acta*. 2003;334:5-23.

13. Wright EM, Royston P. Calculating reference intervals for laboratory measurements.

*Stat Methods Med Res*. 1999;8:93-112.[DOI: 10.1191/096228099672225506]

14. Kronish IM, Cheung YK, Shimbo D, Julian J, Gallagher B, Parsons F, et al. Increasing the precision of hypertension treatment through personalized trials: a pilot study. *J Gen Intern Med*. 2019;34:839-45.[DOI: 10.1007/s11606-019-04831-z]

15. Parati G, Bilo G, Kollias A, Krzesinski J, Peyrard S, Tykarski A, et al. Blood pressure variability: methodological aspects, clinical relevance and practical indications for management—a European Society of Hypertension position paper. *J Hypertens*. 2023;41:527-44.[DOI: 10.1097/HJH.0000000000003363]

16. Chowdhury M, Naeem I, Quan H, Leung AA, Sikdar KC, O'Beirne M, et al. Prediction of hypertension using traditional regression and machine learning models: a systematic review and meta-analysis. *PLoS One*. 2022;17:e0266334.[DOI: 10.1371/journal.pone.0266334]

17. Parati G, Stergiou GS, Bilo G, Kollias A, Pengo M, Ochoa JE, et al. Home blood pressure monitoring: methodology, clinical relevance and practical application: a 2021 position paper by the Working Group on Blood Pressure Monitoring and Cardiovascular Variability of the European Society of Hypertension. *J Hypertens*. 2021;39:1742-67.[DOI: 10.1097/HJH.0000000000002922]

18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-74.[DOI: 10.2307/2529310]

19. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420-8.[DOI: 10.1037/0033-2909.86.2.420]

20. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-63.[DOI: 10.1016/j.jcm.2016.02.012]

21. Bland JM, Altman DG. Statistics notes: measurement error. *BMJ*. 1996;312:1654.[DOI: 10.1136/bmj.312.7047.1654]

22. Rothwell PM. Limitations of the usual blood-pressure hypothesis and importance of variability, instability, and episodic hypertension. *Lancet*. 2010;375:938-48.[DOI: 10.1016/S0140-6736(10)60309-1]

23. Feng Y, Zhang X, Song J, Yang X, Yao J, Liu W, Feng L. Between- and within-person components of blood pressure: a two-component distribution model and the informational limits of individualised correction. Submitted to Physiological Measurement.