

Accuracy in Labeling: A Study on the Composition and Nutritional Grading of Freshly Made Beverages in China

Zhuo Sun^{1,2,†}, Xinyi Lu^{3,†}, Yichao Wang⁴, Zhengyuan Wang², Wenqing Ma²,
Liping Shen², Shupeng Mai², Qi Song², Wei Lu², Zehuan Shi^{2,*}, Jiajie Zang^{2,*}

¹School of Public Health, Fudan University, 200336 Shanghai, China

²Division of Health Risk Factors Monitoring and Control, Shanghai Municipal Center for Disease Control and Prevention, 200336 Shanghai, China

³School of Public Health, Shanghai University of Traditional Chinese Medicine, 201203 Shanghai, China

⁴College of Medical Technology, Shanghai University of Medicine & Health Sciences, 200332 Shanghai, China

*Correspondence: shizehuan@scdc.sh.cn (Zehuan Shi); zangjiajie@scdc.sh.cn (Jiajie Zang)

†These authors contributed equally.

Academic Editor: Torsten Bohn

Submitted: 9 January 2026 Revised: 27 July 2026 Accepted: 30 July 2026 Published: 20 October 2026

Abstract

Background: Freshly made beverages (FMBs) have become an increasingly important source of sugar and fat intake in China. However, the on-site preparation and product customization of these beverages make conventional tolerance frameworks for nutrition labeling difficult to implement. Thus, this study aimed to assess the accuracy of Nutri-Choice, a voluntary graded nutrition labeling system piloted in Shanghai, by comparing brand-registered nutrient values with laboratory-measured values under routine retail conditions. **Methods:** Three major chain brands involved in the initial Nutri-Choice pilot were included. The system categorizes beverages into four color-coded grades (A to D) according to predefined thresholds for non-milk extrinsic sugars (NMESs, defined as total sugars minus lactose), saturated fatty acids (SFAs), trans fatty acids (TFAs), and non-sugar sweeteners. In total, 122 beverages were selected through stratified random sampling and analyzed for NMESs, SFAs, and TFAs. Laboratory-measured values were compared with brand-registered values using relative error analysis, weighted agreement between actual grading and brand-registered grading, Cohen's kappa statistics, and Bland–Altman analysis. **Results:** Among the 122 samples, pure tea accounted for 9.02%, fruit tea and fruit juice for 28.69%, and milk coffee and milk tea for 62.29%. The median contents of NMESs, SFAs, and TFAs were 3.63 [1.31, 5.72], 0.90 [0.00, 2.10], and 0.00 [0.00, 0.00] g/100 mL, respectively. Grade B beverages were the most frequent (46.72%), followed by Grade C (25.41%), Grade D (22.95%), and Grade A (4.92%). The NMES concentrations increased from Grades A to C, whereas those in Grade D were mainly determined by SFA levels. Weighted relative error analysis showed that 85.63% of the NMES measurements, 79.38% of the SFA measurements, and 98.75% of the TFA measurements satisfied the $\leq 120\%$ claimed-value criterion. Bland–Altman analysis revealed a significant negative bias for NMESs, suggesting that laboratory-measured values were, on average, 0.82 g/100 mL lower than brand-registered values. No systematic bias was detected for SFAs or TFAs. The initial weighted grade agreement was 81.88% (Kappa = 0.688), which increased to 88.68% (Kappa = 0.812) after applying a 120% adjustment coefficient. **Conclusions:** Nutri-Choice provides a practical framework for graded nutrition labeling of FMBs, although the credibility of this framework depends on routine market sampling, laboratory verification, and evidence-based tolerance approaches. The observed variation in grading agreement among brands indicates the need for further assessment of differentiated tolerance.

Keywords: beverages; sugar-sweetened beverages; food labeling; nutritive value; food analysis

1. Background

With the rapid expansion of chain freshly made beverage (FMB) enterprises and the growing online-to-offline (O2O) food delivery industry, China's FMB market has shown substantial growth, with its scale expected to reach USD 28.0 billion by 2025 [1]. The consumption of FMBs, including freshly made coffee, Asian-style beverages, and fruit & milk teas, has continued to increase and has gradually become an important source of free sugar intake among Chinese residents [2,3]. This trend creates considerable public health challenges. A robust body of evidence has linked habitual intake of sugar-sweetened beverages with weight gain and a higher risk of type 2 diabetes, cardiovas-

cular diseases, and certain cancers [4]. An umbrella review published in 2023 further confirmed that sugar-sweetened beverage consumption is significantly associated with increased risks of gout, metabolic syndrome, obesity, type 2 diabetes, hypertension, coronary heart disease, stroke, cancer, and all-cause mortality [5], underscoring the importance of reducing their intake for public health. Evidence from dietary guidelines shows that sugar-sweetened beverages (SSBs) account for 42.1% of free sugar intake among urban residents, and that 18.9% of primary and secondary school students frequently consume SSBs [6,7]. FMBs involve unique risks for two main reasons. First, many products contain high amounts of sugar and fat. For example,

93.3% of products claiming “no added sugar” were reported to contain more than 0.5 g/100 g of sugar, while the median fat and trans fatty acids (TFAs) contents in beverage samples collected in Shanghai in 2021 were 2.48 [1.80, 3.15] g/100 g and 0.721 [0, 2.22] g/100 g, respectively [8]. Second, regulatory gaps still exist. China’s current food safety standard, GB 28050-2025, does not require sugar or fat labeling for FMBs [9], leaving consumers with limited information on the actual sugar and fat contents of these beverages.

To address this problem, Shanghai has introduced the voluntary Nutri-Choice labeling system for FMBs. This system adapts Singapore’s *Nutri-Grade* framework and classifies beverages into four grades, from A to D, according to non-milk extrinsic sugars (NMESs), saturated fatty acids (SFAs), TFAs, and non-sugar sweeteners (NSSs) [10,11]. The first batch of pilot FMB enterprises included three well-known FMB chain brands in China. Nevertheless, quality-control difficulties remain. Variations in ingredient batches (e.g., syrups and fresh fruits), together with manual preparation errors, may cause marked fluctuations in nutrient content [8]. For products containing fruit or cream toppings, deviations between actual and labeled nutrient values may exceed 20%. Even among chain brands, direct-sale and franchised stores may vary in their adherence to standard operating procedures (SOPs) during beverage preparation [12].

The Chinese National Food Safety Standard General Rules for Nutrition Labeling of Prepackaged Foods (GB 28050-2025), requires declared values for encouraged nutrients (e.g., protein, fiber), to be at least 80% of labelled values, whereas restricted components (e.g., fat, SFA, TFA, NMES, sodium), should not exceed 120% of labelled values, with vitamins A/D permitted 80–180% ranges because of greater variability [9]. However, FMBs are not covered by mandatory nutrition labeling under GB 28050-2025 because foods prepared and sold on site are exempt from these requirements [9]. Most previous beverage studies have focused on nutrient profiling, sugar surveillance, or the design of front-of-package labeling (FOPL) systems, while few have examined whether nutrient grades submitted by manufacturers remain accurate when beverages are prepared and purchased under routine retail conditions. This distinction is important because FMBs are made at the point of sale and may be influenced by ingredient batching, manual portioning, toppings, ice content, and store-level compliance with SOPs. Thus, evidence is still limited regarding how the accuracy of nutrition labeling should be monitored for beverages prepared on site.

This study examined the accuracy of Nutri-Choice labeling among the first three pilot FMB brands through randomized sampling and laboratory testing of NMESs, SFAs, and TFAs. The objectives were to: (1) evaluate the agreement between brand-registered and laboratory-measured Nutri-Choice grades; (2) identify differences be-

tween brand-registered nutrient values and market-sampled laboratory measurements; and (3) explore brand-stratified adjustment coefficients associated with improved grading agreement to support refinement of the grading system.

2. Methods

2.1 Data Source

Before the pilot project began in March 2024, the three participating companies submitted their original registration materials to the project initiator, the Shanghai Municipal Center for Disease Control and Prevention (SCDC). These materials included nutritional data and testing reports for all beverage products or ingredients. All compliance testing reports were issued by qualified third-party testing institutions with China National Accreditation Service for Conformity Assessment (CNAS) authorized detection qualifications.

Three national chain FMB manufacturers involved in the first Nutri-Choice pilot program were included and anonymized as Brands X, Y, and Z. Before the program was implemented, each manufacturer provided complete nutritional datasets to SCDC, including: (1) product formulation specifications for all beverage categories; (2) compliance testing reports confirming nutrient composition, issued by a CNAS-accredited third-party laboratory; and (3) SOPs for ingredient preparation and quality control. Nutrient levels obtained from brand registration were expressed as Nutrient_i registered by the brand.

The three brands varied in their preparation procedures and operational characteristics, as described below. (1) Brand X represented standardized production, with automated tea-making machines using preset programs and maintaining a $\leq 5\%$ coefficient of variation in nutrient output. (2) Brand Y represented predominantly handmade production, with more than 80% of preparation steps completed manually and therefore requiring barista skill calibration. (3) Brand Z represented hybrid production, combining partial automation with manual assembly of fruit or cream toppings, with approximately 15–20% of steps performed manually.

2.2 Sample Size

To maintain both efficiency and representativeness, stratified random sampling was used as the sampling strategy in this study. For the estimation of population proportions, the sample size was determined using Eqn. 1:

$$n = \frac{z^2 \times p(1 - p)}{E^2} \quad (1)$$

In this study, the confidence level was defined as 95% ($z = 1.96$), the expected compliance rate was set at 50% ($p = 50\%$), and the margin of error was set at 9% ($E = 0.09$). According to Eqn. 1, the required sample size was 119. The

Table 1. Sampling allocation by brand and production modality.

Brand	Total stores in Shanghai	Selected stores	Registered SKUs	Sampling weight (%)	Target sampled SKUs	Rationale
Brand X	84	4	473	5	24	Machine standardization was expected to minimize nutrient variability.
Brand Y	27	13	617	12	74	High operator dependency required greater verification intensity.
Brand Z	122	4	233	10	24	Moderate manual intervention required intermediate monitoring intensity.
Total	233	21	1323	—	122	—

Note: (1) SKU, stock-keeping unit; (2) Sampling weight refers to the proportion of registered SKUs selected within each brand; (3) Target sampled SKUs represent the number of beverage samples collected and analyzed; (4) Store totals include both direct-sale and franchised outlets.

final sample size was then adjusted on the basis of practical factors, including workforce availability, cost limitations, and the distribution of stock-keeping units (SKUs) across brands.

Based on the sampling design of Liu et al. [13], in which 708 stratified beverage samples were collected for free sugar monitoring in China, and considering the 8236 beverage products reported by Li et al. [14], the sampling proportion was estimated to be approximately 9%. To balance representativeness and operational feasibility, sampling weights of 5%, 12%, and 10% were assigned to Brands X, Y, and Z, respectively, according to their production modalities and the rationale presented in Table 1. Therefore, the final target sample sizes were 24 samples from Brand X, corresponding to 5% of 473 registered SKUs; 74 samples from Brand Y, corresponding to 12% of 617 registered SKUs; and 24 samples from Brand Z, corresponding to 10% of 233 registered SKUs. The final sample size was thus 122 beverages.

2.3 Randomization Process

Beverage sample randomization and sampling-store selection were conducted using R v4.0.5 (Version 4.0.5; R Foundation for Statistical Computing, Vienna, Austria; <https://www.R-project.org/>). Unique identification numbers were assigned to each product from a total pool of 1323 registered beverage SKUs with complete nutritional information. The order of SKU selection for sampling was determined by a random number sequence generated in R.

All 233 retail outlets from the three participating brands (X, Y, Z) in Shanghai were incorporated into the sampling framework. To balance practical constraints, including manpower and cost, with statistical representativeness, stores were selected as follows: 4 stores from Brand X (4.76% selection probability per store); 13 stores from Brand Y (48.15% selection probability per store); and 4 stores from Brand Z (3.28% selection probability per store).

Samples were purchased from each selected store according to their randomized order. The sampling sequence used strict numerical allocation: Store 1 collected sam-

ples #1–6, Store 2 collected #7–12, etc. For unavailable products, contingency procedures were followed, and replacements were selected using subsequent random numbers from the predetermined sequence. The detailed workflow is shown in Fig. 1.

2.4 Sample Collection Protocol

During each sampling session, beverage samples were collected by at least two qualified professionals. A single-blind approach was used for store personnel, ensuring that beverage producers were unaware of the study objectives and that each purchase was treated as an ordinary consumer transaction. After collection, samples were retained in sealed packaging to reduce the possibility of external contamination and were delivered to the laboratory for analysis within two hours. During transport, samples were kept at temperatures close to their original storage conditions to maintain sample integrity.

2.5 Sample Classification Criteria

In accordance with the Chinese beverage classification standard GB/T 10789-2015 [15], criteria for sweetness, temperature, cup type, and beverage type were standardized, as presented in Table 2 (Ref. [16,17,18]).

2.6 Laboratory Testing

The 122 collected beverages were sent to a testing institution with China Metrology Accreditation (CMA) and CNAS qualifications. The testing items and analytical methods were as follows. First, NMESs, SFAs, and TFAs levels in the samples were measured. For NMESs, total sugar content, including lactose, fructose, glucose, sucrose, and maltose, was determined using the high-performance liquid chromatography (HPLC) method specified in GB 5009.8-2023 [19]. NMESs were calculated by subtracting lactose from total sugar. SFAs were measured using the normalization method specified in GB 5009.168-2016 [20], whereas TFAs were determined using the gas chromatographic method described in GB 5009.257-2016 [21]. All analyses were performed according to national stan-

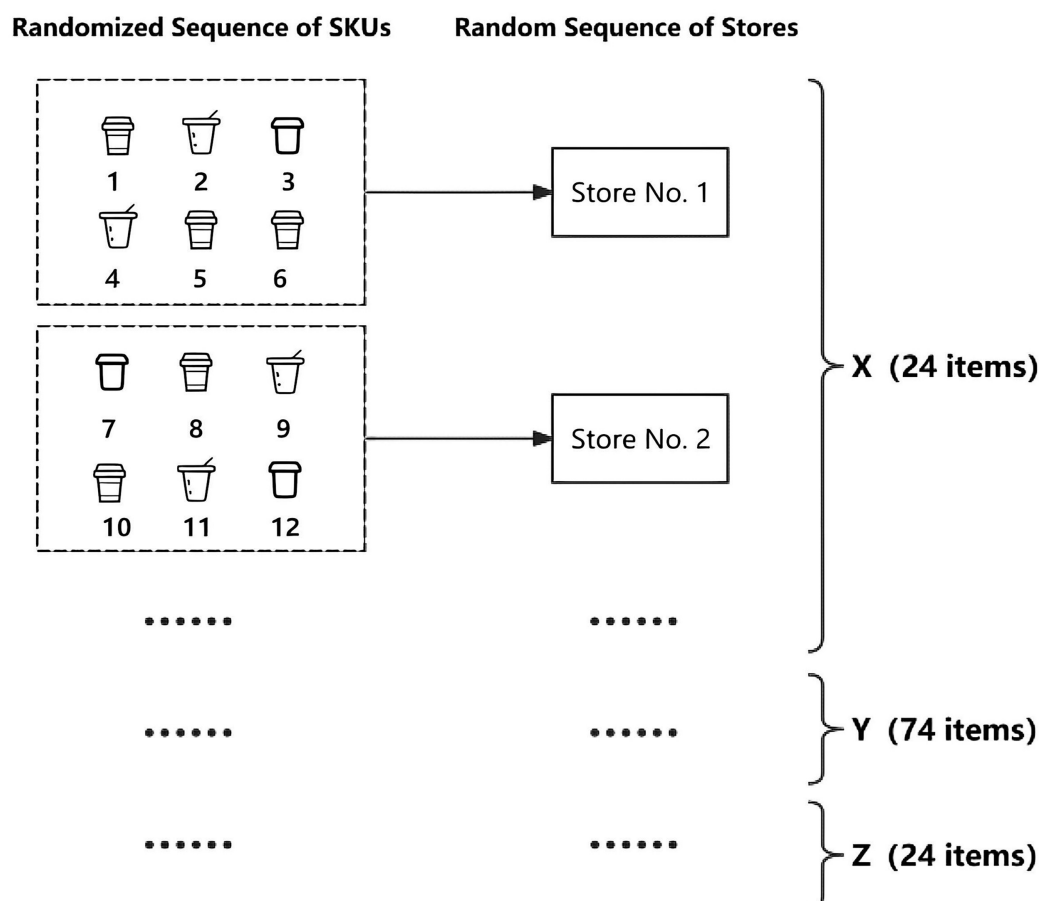


Fig. 1. Randomization and allocation workflow for beverage sampling. Note: (1) Registered SKUs and eligible stores were randomized within each brand. Selected SKUs were then sequentially allocated to selected stores until the target sample sizes were reached for Brand X ($n = 24$), Brand Y ($n = 74$), and Brand Z ($n = 24$); (2) SKU, stock-keeping unit.

dard protocols. Density was used to convert laboratory-measured values and standardize the units from g/100 g to g/100 mL. Nutrient values obtained by laboratory analysis were defined as Nutrient_i from actual testing. NSS status was not measured in the laboratory but was determined according to whether NSSs were declared in the beverage formula provided by the manufacturer.

The testing protocol required samples containing components such as cheese, milk foam, crushed nuts, cream, or other added ingredients to be thoroughly mixed before analysis. If inedible fruits, such as flattened lemons or kumquats, were present in the sample, these components were removed before testing. Ice-containing samples were fully blended to ensure accurate measurement.

2.7 Classification Criteria for Nutri-Choice

Nutri-Choice applies a graded classification system with corresponding standards and implementation rules. Classification is based on a threshold-based nutrient assessment model, in which thresholds are specified for NMESs, SFAs, TFAs, and NSSs, as shown in Fig. 2 and Table 3. Further information on the scope, grading criteria, label pre-

sentation, and implementation of Nutri-Choice is available in the official Specification for Nutri-Choice Labeling of Fresh Made Beverages [22]. Nutrient-profile labeling systems generally prioritize nutrients associated with the most pressing public health concerns, which currently include sugars, fats, and sodium. Because sodium is not a primary nutritional concern for freshly made beverages, the Nutri-Choice criteria focus on NMESs, SFAs, TFAs, and NSSs. At present, Nutri-Choice is applied exclusively to freshly made beverages and has not been extended to other food categories.

Nutri-Choice labels are grouped into four grades, from A to D, reflecting progressively lower recommendation levels. Grade A indicates the most recommended option, while Grade D indicates the least recommended option. The four grades are represented by dark green, light green, yellow, and red, respectively. Grade A beverages are displayed in dark green and correspond to the lowest NMESs, SFAs, and TFAs levels and the absence of NSSs; Grade D beverages are displayed in red and correspond to the highest NMESs, SFAs, and TFAs levels. In this study, NMESs were calculated as total sugars minus lactose. This operational mea-

Table 2. Criteria and definitions used for sample classification.

Drink option	Standardized classification	Another name/definition and source
Sweetness	Full sugar	Standard sugar, Standard sweet, normal sugar, normal sweet
	Less sugar	Less sugar, less sweet, 3/4 sugar
	Half sugar	Half sugar, 1/2 sugar
	Quarter sugar	Slightly sweet, slightly sugar, 1/4 sugar
	No sugar	Sugar-free, 0 sugar
Temperature	Cold	Cold, ice-free, ice, standard ice, less ice, more ice, ice sand
	Hot	Room temperature, warm, hot, scalding, piping hot
Cup size	Regular	350–500 mL
	Large	600–650 mL
	Jumbo	900 mL
Drink types	Pure tea	Tea, as defined by the national standard GB/T 21733-2008 [16], is a liquid beverage that is made from the water extract of tea or its concentrated liquid, tea powder, etc., which maintains the original flavor of tea juice.
	Fruit Tea	Juice tea drinks and fruity tea drinks, as defined by the national standard GB/T 21733-2008 [16], is a liquid beverage made from tea ingredients, adding juice, edible fruit essence, etc.
	Fruit juice	Fruit and vegetable juice (pulp) drinks, as defined by the national standard GB/T31121-2014 [17], is a liquid beverage made from fresh or concentrated fruit and vegetable juice (or pulp).
	Milk coffee	Coffee drinks, as defined by the national standard GB/T30767-2014 [18], is a liquid beverage made from coffee beans/ coffee products (ground coffee powder, coffee extract or its concentrate, instant coffee, etc.).
	Milk tea	Milk tea drinks and milk-flavor tea drinks, as defined by the national standard GB/T 21733-2008 [16], is a liquid beverage made from tea ingredients, with milk or dairy products.

Table 3. Nutrient thresholds for Nutri-Choice Grades A to D.

Grade	Specific requirements (g/100 mL)
A	NMESs ≤ 0.5 , SFAs ≤ 0.75 , TFAs ≤ 0.3 and no NSSs
B	$0.5 < \text{NMESs} \leq 5$, $0.75 < \text{SFAs} < 2.5$, TFAs ≤ 0.3
C	$5 \leq \text{NMESs} \leq 11.5$, $0.75 \leq \text{SFAs} \leq 2.5$, TFAs > 0.3
D	NMESs > 11.5 , SFAs > 2.5 , TFAs > 0.3

Note: NMESs, Non-milk Extrinsic Sugars, namely total sugar (g/100 mL) – lactose (g/100 mL); SFAs, Saturated Fatty Acids; TFAs, Trans Fatty Acids; NSSs, Non-Sugar Sweeteners.

sure was adapted from the Singapore Nutri-Grade framework [23] and is consistent with the concept of NMESs used in previous literature [24]. By excluding naturally occurring milk sugars, NMESs reduce the influence of dairy-derived sugars on beverage grading and facilitate comparisons across beverages with different milk contents. The final overall beverage grade is determined by the lowest grade assigned to any individual nutrient component. An example of Nutri-Choice logo use is presented in Fig. 3.

2.8 The Judgment of the Consistent Classification

(1) Relative Error (RE) was determined as the difference between the laboratory-measured nutrient value and the brand-registered value, divided by the brand-registered value. The RE formula is:

$$RE = \frac{\text{Nutrient}_i \text{ from actual testing} - \text{Nutrient}_i \text{ registered by the brand}}{\text{Nutrient}_i \text{ registered by the brand}} \times 100\%$$

For samples with a registered value of zero, the denominator in the RE formula was replaced with a very small positive constant (1×10^{-6}) to avoid division by zero. If the measured value was also zero, the RE was recorded as 0%. For restricted nutrients, including NMESs, SFAs, and TFAs, $RE \leq 20\%$ means that the laboratory-measured value falls within the allowable tolerance range and does not exceed 120% of the brand-registered value. A negative RE shows that the laboratory-measured value is lower than the brand-registered value. Conversely, $RE > 20\%$ indicates

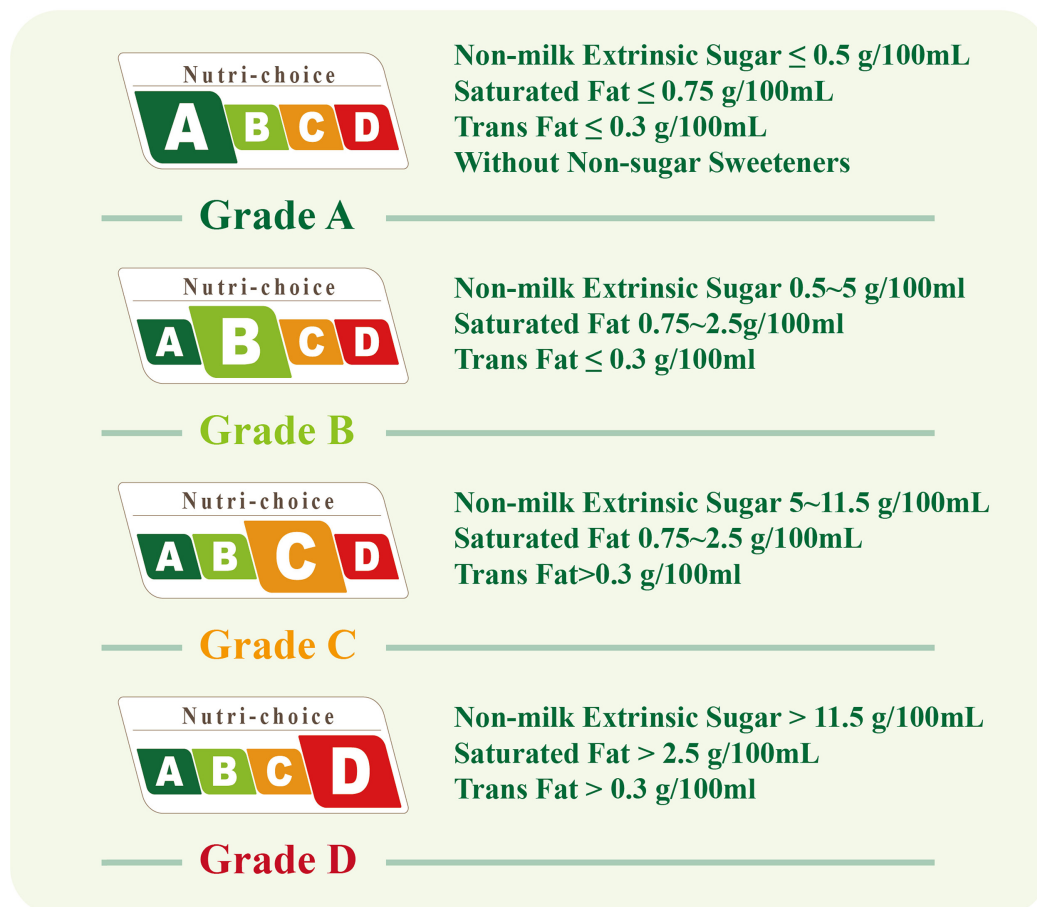


Fig. 2. Overview of the Nutri-Choice grading system and nutrient thresholds.

that the laboratory-measured value is beyond the permitted tolerance range and exceeds 120% of the brand-registered value.

(2) Actual grading (AG) refers to the Nutri-Choice grade calculated from laboratory-measured nutrient values. Brand-registered grading (BG) refers to the Nutri-Choice classification derived from brand-registered nutrient values submitted by the participating brands.

Compliance rate was calculated as the proportion of samples where actual grading (AG, derived from laboratory-measured nutrient values) matched brand-registered grading (BG, derived from brand-registered nutrition values). Regulatory standards previously permitted unidirectional tolerance for restricted nutrients: the measured value may be equal to or lower than the labeled value, but must not exceed 120% of the labeled value [9,25,26]. In the context of Nutri-Choice grading, this means that the AG can be equal to or better than the BG, but cannot be worse. For example, a BG-classified “C Grade” product could achieve AG “B Grade” (improved) but not “D Grade” (worse). This metric evaluates alignment between laboratory-verified classifications and brand-declared categorizations, considering consistency when AG meets or exceeds BG.

The formula for Compliance rate is:

$$\text{Compliance rate (\%)} = \frac{N_{\text{AG-BG matches}}}{N_{\text{total}}} \times 100\%$$

2.9 Statistical Analysis

Experimental data were systematically recorded in Excel. Laboratory-measured values were compared with brand-registered values to determine Nutri-Choice grades and compliance rates. Normality was assessed using the Shapiro-Wilk test. Because the dataset did not follow a normal distribution, continuous variables were reported as medians with interquartile ranges. Group differences were analyzed using the Kruskal-Wallis test, followed by post hoc pairwise comparisons with Bonferroni correction.

Agreement between AG and BG was evaluated using Cohen’s kappa statistic. To consider the unequal sampling fractions across brands, brand-level sampling weights were incorporated into the weighted kappa analysis to adjust for the unbalanced sample sizes among the three brands. The kappa coefficient ranges from -1 to 1 , with higher absolute values indicating stronger agreement, whereas negative values indicating agreement below that expected by chance. Interpretation was based on conventional thresholds: $[-1,$

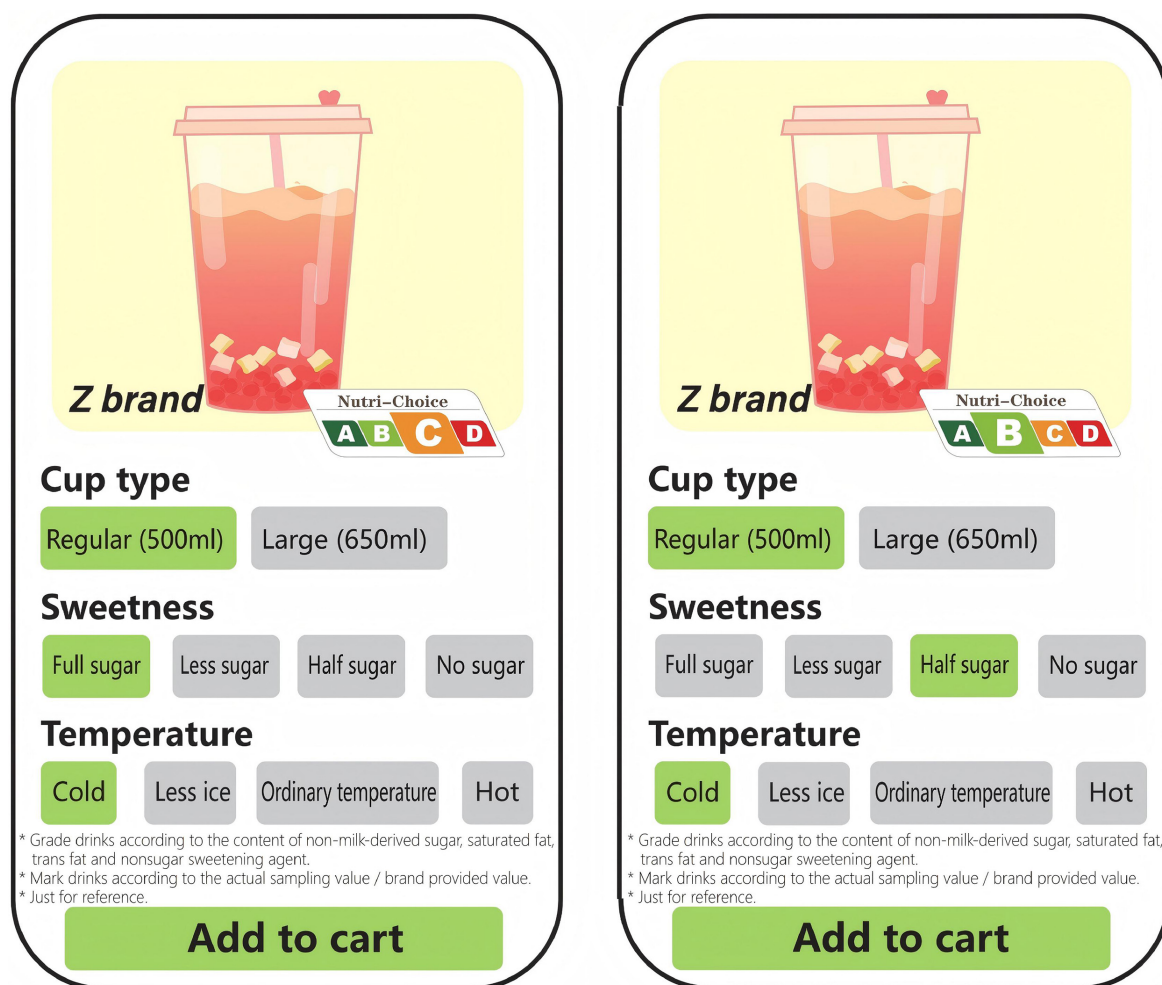


Fig. 3. Example of Nutri-Choice grade presentation in an online ordering interface.

0) = poor agreement, 0 = random agreement, (0, 0.2] = slight agreement, (0.2, 0.4] = fair agreement, (0.4, 0.6] = moderate agreement, (0.6, 0.8] = high agreement, and (0.8, 1] = almost perfect agreement [27].

Bland-Altman analysis was applied to evaluate agreement between laboratory-measured nutrient values and brand-registered specifications for NMESs, SFAs, and TFAs. The differences between paired measurements were plotted against the mean nutrient concentration calculated from the two measurements. Analytical thresholds were marked using horizontal reference lines: the central line indicated the mean systematic bias, while the upper and lower boundaries represented the 95% limits of agreement (LoA), calculated as the mean discrepancy ± 1.96 standard deviations of the observed differences.

3. Results

3.1 Basic Characteristics and Nutritional Distribution of Laboratory-Measured Samples

The median contents of NMESs, SFAs and TFAs were 3.63 [1.31, 5.72] g/100 mL, 0.90 [0.00, 2.10] g/100 mL, and 0.00 [0.00, 0.00] g/100 mL, respectively (Table 4). Grade B

beverages were the most common category, comprising 57 samples (46.72%), followed by Grade C with 31 samples (25.41%), Grade D with 28 samples (22.95%), and Grade A with 6 samples (4.92%).

In subgroup comparisons, Brand X showed significantly lower NMESs levels than Brands Y and Z ($p < 0.05$), while no significant difference was observed between Brands X and Z for NMESs. Brand Z showed notably lower SFAs levels than the other two brands. The Nutri-Choice system effectively separated NMESs content across Grades A, B, and C. However, Grade C beverages had higher NMESs levels than Grade D beverages, whereas Grade D beverages showed the highest SFAs levels. This pattern indicates that Grade D classification in these samples was mainly driven by SFAs levels rather than by NMESs levels alone. No significant differences in NMESs or SFAs were detected among temperature or cup-size groups.

According to declared sweetness, NMESs levels were highest in full-sugar beverages and lowest in beverages labeled as no extra sugar. However, the upper quartile of NMESs in the no extra sugar group reached 2.94 g/100 mL, showing that measurable NMESs were still detected in

Table 4. Distribution of laboratory-measured NMESs, SFAs and TFAs in sampled samples (N = 122).

Nutrient (g/100 mL)			NMESs	SFAs	TFAs
Classification	Subunit	N (%)	Median [Q1, Q3]	Median [Q1, Q3]	Median [Q1, Q3]
Total		122 (100.00)	3.63 [1.31, 5.72]	0.90 [0.00, 2.10]	0.00 [0.00, 0.00] ^a
Brand	X brand	24 (19.67)	2.85 [0.00, 4.31] ^a	1.56 [0.72, 1.97] ^a	0.00 [0.00, 0.00] ^a
	Y brand	74 (60.66)	4.12 [1.62, 5.93] ^b	1.21 [0.00, 2.59] ^a	0.00 [0.00, 0.00] ^a
	Z brand	24 (19.67)	4.53 [1.30, 7.75] ^{ab}	0.00 [0.00, 0.68] ^b	0.00 [0.00, 0.00] ^a
Nutri-Choice classification	Grade A	6 (4.92)	0.00 [0.00, 0.00] ^a	0.00 [0.00, 0.35] ^a	0.00 [0.00, 0.00] ^a
	Grade B	57 (46.72)	2.43 [0.00, 3.53] ^a	1.07 [0.00, 1.67] ^a	0.00 [0.00, 0.00] ^a
	Grade C	31 (25.41)	6.58 [5.56, 7.94] ^b	0.00 [0.00, 0.70] ^a	0.00 [0.00, 0.00] ^a
	Grade D	28 (22.95)	4.50 [2.95, 7.53] ^c	2.68 [2.09, 3.14] ^b	0.00 [0.00, 0.00] ^a
Declared sweetness	Full sugar	21 (17.21)	5.90 [5.16, 8.31] ^c	0.00 [0.00, 1.77] ^a	0.00 [0.00, 0.00] ^a
	Less sugar	23 (18.85)	4.47 [3.64, 5.61] ^{bc}	0.81 [0.00, 2.13] ^a	0.00 [0.00, 0.00] ^a
	Half sugar	16 (13.11)	3.41 [3.03, 4.43] ^{ab}	0.91 [0.00, 2.31] ^a	0.00 [0.00, 0.00] ^a
	Quarter sugar	24 (19.67)	2.83 [1.23, 6.25] ^{ab}	1.11 [0.00, 2.42] ^a	0.00 [0.00, 0.00] ^a
	No extra sugar	38 (31.15)	0.00 [0.00, 2.94] ^a	1.21 [0.00, 2.32] ^a	0.00 [0.00, 0.00] ^a
Temperature	Cold	62 (50.82)	3.54 [1.38, 5.92] ^a	0.81 [0.00, 2.08] ^a	0.00 [0.00, 0.00] ^a
	Hot	60 (49.18)	3.63 [1.23, 5.60] ^a	1.21 [0.00, 2.21] ^a	0.00 [0.00, 0.00] ^a
Cup size	Regular	67 (54.92)	3.24 [1.21, 5.31] ^a	0.91 [0.00, 2.08] ^a	0.00 [0.00, 0.00] ^a
	Large & Jumbo	55 (45.08)	4.22 [1.40, 6.45] ^a	0.81 [0.00, 2.19] ^a	0.00 [0.00, 0.00] ^a
Drink types	Pure tea	11 (9.02)	3.16 [1.72, 4.47] ^a	0.00 [0.00, 0.00] ^a	0.00 [0.00, 0.00] ^a
	Fruit Tea & Fruit juice	35 (28.69)	7.59 [3.36, 9.27] ^b	0.00 [0.00, 0.00] ^a	0.00 [0.00, 0.00] ^a
	Milk coffee & Milk tea	76 (62.29)	3.04 [0.00, 4.81] ^a	1.80 [1.01, 2.57] ^b	0.00 [0.00, 0.00] ^a

Note: (1) Values are presented as n (%) or median [Q1, Q3]; (2) Overall differences among groups were assessed using the Kruskal-Wallis test, followed by Bonferroni-adjusted post hoc pairwise comparisons; detailed pairwise comparison results are presented in Appendix Table 7; (3) Groups sharing at least one superscript letter are not significantly different, whereas groups with no superscript letters in common differ significantly ($p < 0.05$); (4) NMESs, non-milk extrinsic sugars; SFAs, saturated fatty acids; TFAs, trans fatty acids.

some products with this declaration. By beverage type, fruit tea and fruit juice contained higher NMES levels, whereas milk coffee and milk tea contained higher SFAs levels than other categories. The median TFAs level was 0.00 g/100 mL across all subgroups, and TFAs content did not differ significantly among subgroups.

3.2 Discrepancies Between Registered and Laboratory-Measured Values and the Claimed Compliance Rate

Comparative patterns between laboratory-measured values and brand-registered values were illustrated using slope charts (Appendix Figs. 7,8,9). For NMESs, 74.59% of products showed lower laboratory-measured concentrations than the values registered by the brands. By contrast, no obvious directional pattern was detected between laboratory-measured and brand-registered levels of SFAs and TFAs. The TFAs results were highly consistent with the brand claim of “zero TFA”, with detectable TFAs observed in only two milk tea products (1.64%).

Weighted RE analysis comparing laboratory-measured and brand-registered values showed that 85.63% of NMESs, 79.38% of SFAs, and 98.75% of TFAs measurements had low RE values ($\leq 20\%$) and therefore met

the claimed compliance criterion (Appendix Table 8). Detailed RE subgroup analyses stratified by brand and beverage type are further presented in Appendix Table 8.

3.3 Bland-Altman Agreement Analysis of Differences Between Registered and Laboratory-Measured Values

Bland-Altman analysis was conducted to examine agreement between laboratory-measured and brand-registered values for NMESs, SFAs, and TFAs. The corresponding visual results are presented in Figs. 4,5,6. For NMESs, the broad 95% limits of agreement (± 1.96 SD) suggested a certain degree of random variability, with 4.92% of samples (6/122) located outside these limits. SFAs demonstrated narrower agreement, with only 3.27% of samples (4/122) close to or beyond the limits. In comparison, TFAs demonstrated strong consistency, with data points concentrated around the mean difference line and showing very limited dispersion, indicating close agreement between measured and registered values.

Bland-Altman analysis further identified a significant systematic bias for NMESs, with an overall mean difference of -0.82 g/100 mL (95% confidence interval (CI): $-1.19, -0.46$). This result shows that laboratory-measured NMESs values were consistently lower than the corre-

Bias: -0.8224 (95% CI: -1.1878, -0.4571)
Upper LoA: 3.1725 (95% CI: 2.5397, 3.8053)
Lower LoA: -4.8174 (95% CI: -5.4501, -4.1846)

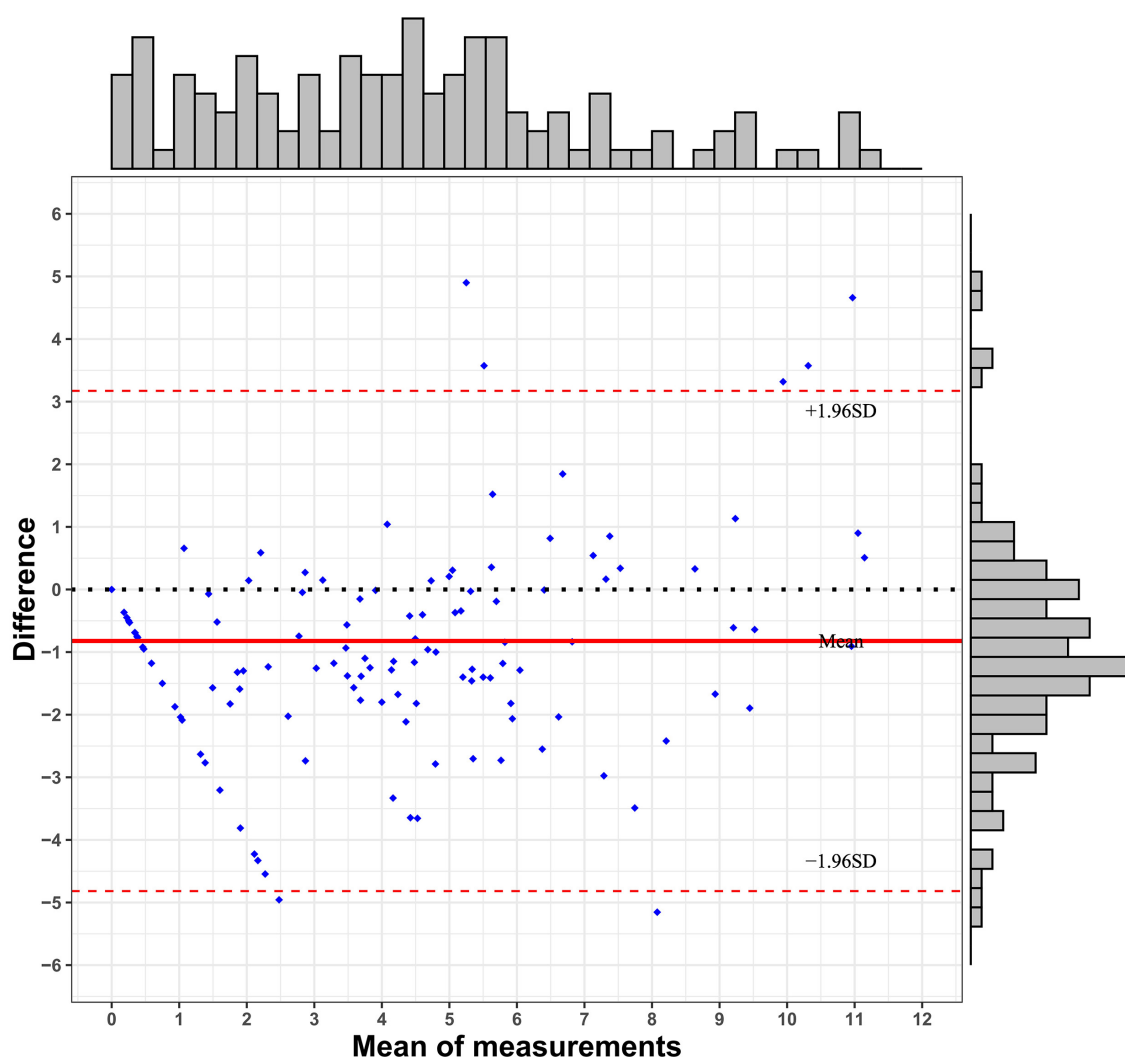


Fig. 4. Bland-Altman plot comparing brand-registered and laboratory-measured NMESs values. Note: (1) Each blue dot represents one beverage sample; (2) X-axis: Mean of the laboratory-measured and brand-registered NMESs values for each sample (g/100 mL); (3) Y-axis: Difference in NMESs values, calculated as laboratory-measured minus brand-registered (g/100 mL); (4) The upper and lower red dashed lines indicate the 95% limits of agreement, and the central red solid line indicates the mean difference.

sponding brand-registered values. However, no systematic bias was detected for SFAs, with a mean difference of -0.06 g/100 mL (95% CI: $-0.20, 0.08$), or for TFAs, with a mean difference of 0.00 g/100 mL (95% CI: $-0.01, 0.02$). Brands X and Y showed clear negative biases, with mean differences of -1.01 g/100 mL (95% CI: $-1.28, -0.74$) and -1.02 g/100 mL (95% CI: $-1.52, -0.52$), respectively, while Brand Z showed no significant bias, with a mean difference of -0.03 g/100 mL (95% CI: $-1.05, 0.98$) (Table 5; Appendix Fig. 10,11,12). These results indicate that the observed deviation in NMESs may have been caused by voluntary sugar reduction and product reformulation by Brands X and Y rather than by methodological variation between laboratories.

3.4 Compliance Rate of AG and BG

The weighted compliance rate reached 81.88%, with Kappa = 0.688 (95% CI: 0.594, 0.783, $p < 0.001$). Brand-specific compliance varied: Brand X showed the highest compliance at 95.83%, Kappa = 0.841 (95% CI: 0.541, 1.000, $p < 0.001$), followed by Brand Z at 75.00%, Kappa = 0.611 (95% CI: 0.350, 0.872, $p < 0.001$), and Brand Y at 72.97%, Kappa = 0.598 (95% CI: 0.465, 0.731, $p < 0.001$) (Table 6). Since multiplying the registered value by a defined coefficient may increase the compliance rate, an additional analysis was conducted to examine improvement in compliance by applying coefficient multipliers (105% to 130%) to brand-registered nutrient values. With an additional coefficient of 120%, the overall Kappa value in-

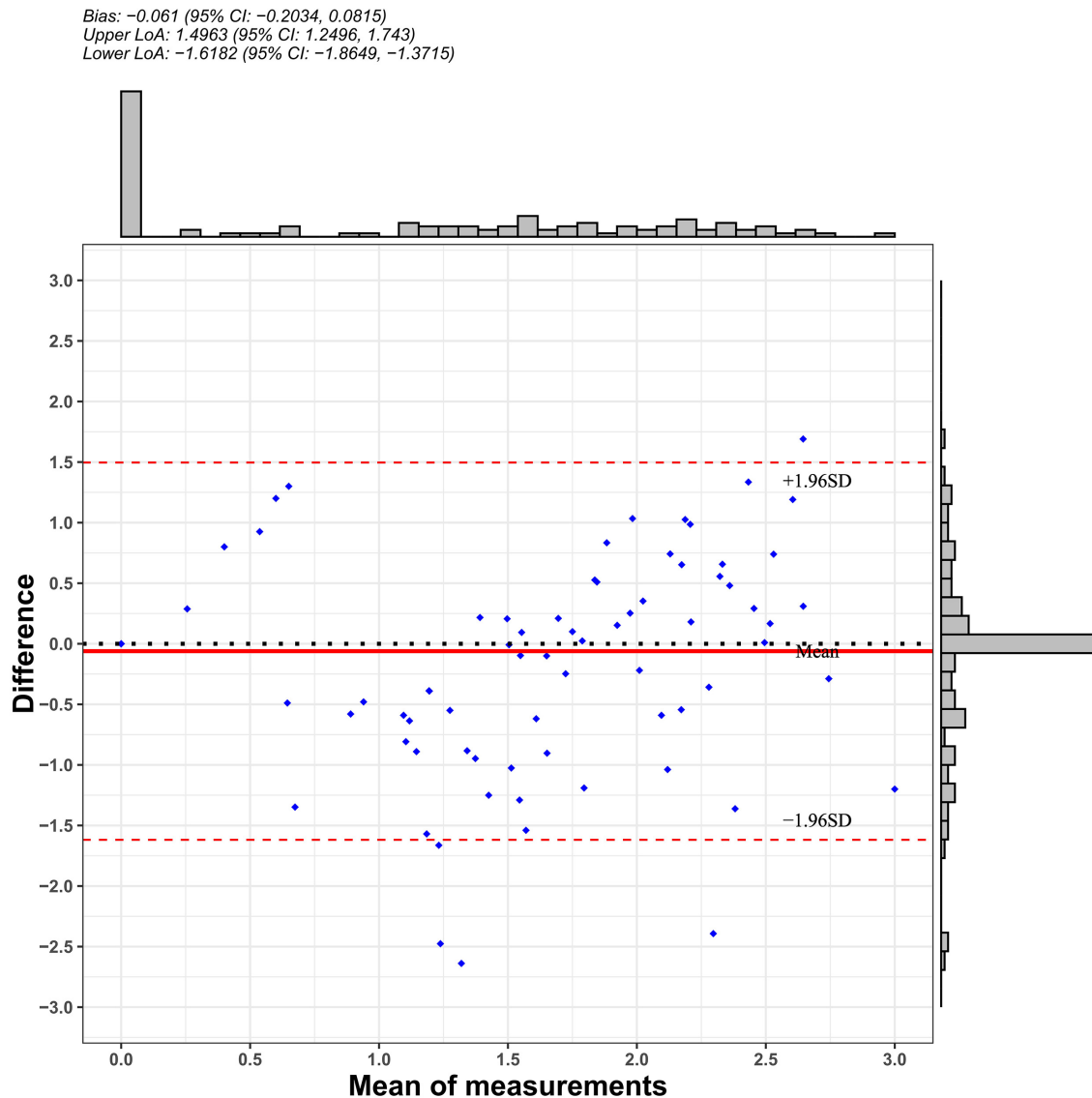


Fig. 5. Bland-Altman plot comparing brand-registered and laboratory-measured SFAs values. Note: (1) Each blue dot represents one beverage sample; (2) X-axis: Mean of the laboratory-measured and brand-registered SFAs values for each sample (g/100 mL); (3) Y-axis: Difference in SFAs values, calculated as laboratory-measured minus brand-registered (g/100 mL); (4) The upper and lower red dashed lines indicate the 95% limits of agreement, and the central red solid line indicates the mean difference.

creased to 0.812 (95% CI: 0.736, 0.888, $p < 0.001$), indicating almost perfect agreement (Table 6). Brand-stratified subgroup analysis showed different patterns of compliance improvement: Brand X achieved 95.83% compliance without adjustment, whereas Brands Z and Y achieved kappa values above 0.8 after applying coefficients of 110% and 125%, respectively. These findings suggest clear heterogeneity in grading agreement among the three brands.

4. Discussion

This study systematically compared laboratory-measured nutrient values with brand-registered nutritional information to examine potential compliance issues in beverage nutrient content claims. Several main findings

were obtained. First, China's first Nutri-Choice-compliant FMBs demonstrated high conformity with restricted nutrient thresholds, as 85.63% of NMESs, 79.38% of SFAs, and 98.75% of TFAs measurements did not exceed 120% of their corresponding brand-registered values. Second, brand-registered values were generally well aligned with laboratory-measured results. Third, laboratory-measured NMESs concentrations were consistently approximately 1 g/100 mL lower than the registered values. Subgroup results further indicated that this decrease was more likely related to voluntary formula optimization by manufacturers than to inter-laboratory variability, suggesting effective reformulation efforts aimed at reducing NMESs. Finally, our exploratory brand-stratified analysis identified adjust-

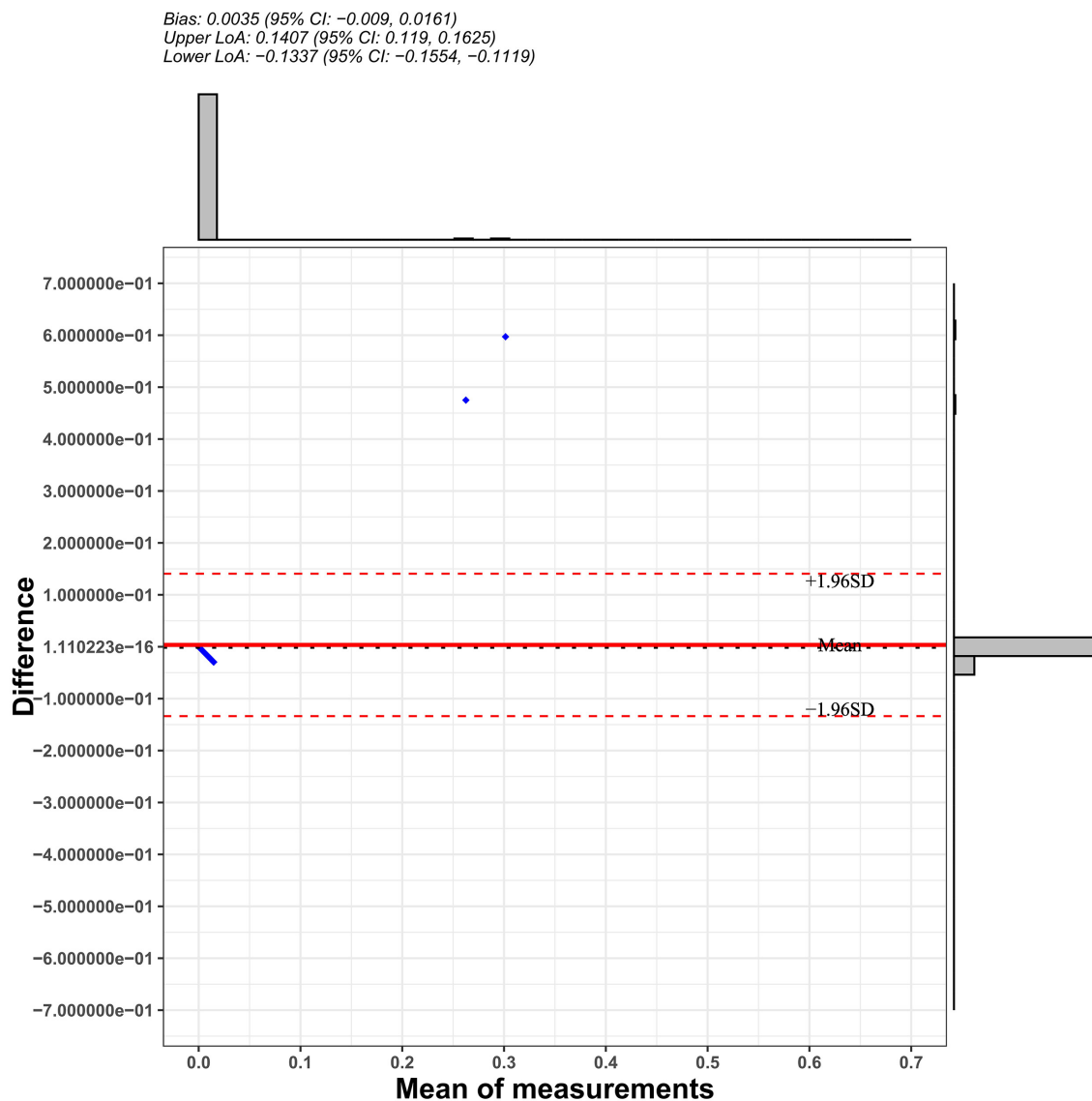


Fig. 6. Bland-Altman plot comparing registered and laboratory-measured TFAs values. Note: (1) Each blue dot represents one beverage sample; (2) X-axis: Mean of the laboratory-measured and brand-registered TFAs values for each sample (g/100 mL); (3) Y-axis: Difference in TFA values, calculated as laboratory-measured minus brand-registered (g/100 mL); (4) The upper and lower red dashed lines indicate the 95% limits of agreement, and the central red solid line indicates the mean difference.

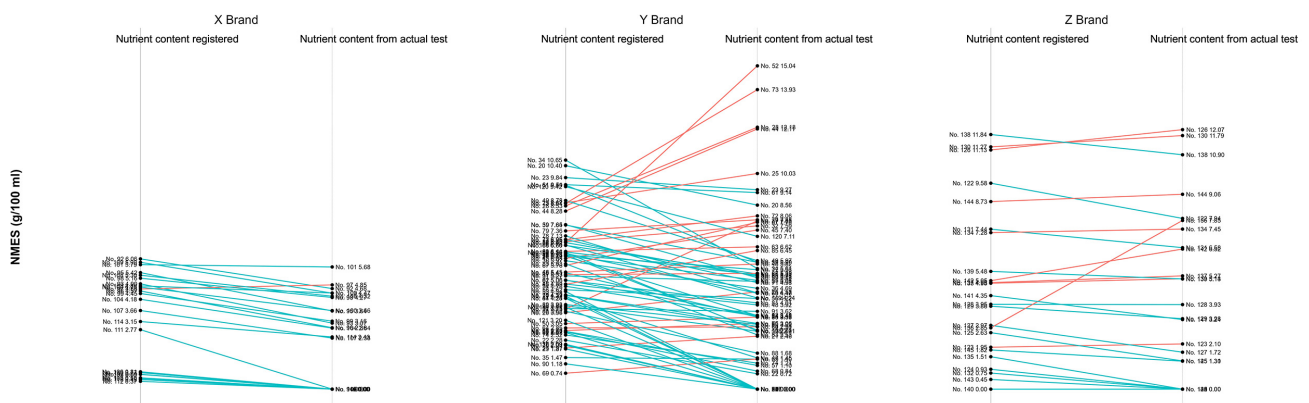


Fig. 7. Differences between Nutrient_i from actual testing and Nutrient_i registered by the brand of various brands of NMEs.

Table 5. Bland-Altman agreement analysis between brand-registered and laboratory-measured nutrient values.

Nutrient	Subgroup	n	Mean difference (95% CI)	Upper LoA (95% CI)	Lower LoA (95% CI)
NMESs (g/100 mL)	Overall	122	−0.82 (−1.19, −0.46)	3.17 (2.54, 3.81)	−4.82 (−5.45, −4.18)
	Z Brand	24	−0.03 (−1.05, 0.98)	4.67 (2.92, 6.43)	−4.74 (−6.50, −2.99)
	Y Brand	74	−1.02 (−1.52, −0.52)	3.22 (2.35, 4.09)	−5.26 (−6.13, −4.39)
	X Brand	24	−1.01 (−1.28, −0.74)	0.25 (−0.22, 0.72)	−2.27 (−2.74, −1.80)
SFAs (g/100 mL)	Overall	122	−0.06 (−0.20, 0.08)	1.50 (1.25, 1.74)	−1.62 (−1.86, −1.37)
	Z Brand	24	−0.02 (−0.28, 0.23)	1.16 (0.72, 1.60)	−1.21 (−1.65, −0.77)
	Y Brand	74	−0.07 (−0.28, 0.15)	1.75 (1.38, 2.12)	−1.89 (−2.26, −1.51)
	X Brand	24	−0.08 (−0.28, 0.12)	0.85 (0.51, 1.20)	−1.01 (−1.35, −0.66)
TFAs (g/100 mL)	Overall	122	0.00 (−0.01, 0.02)	0.14 (0.12, 0.16)	−0.13 (−0.16, −0.11)
	Z Brand	24	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
	Y Brand	74	0.01 (−0.02, 0.03)	0.18 (0.15, 0.22)	−0.17 (−0.21, −0.13)
	X Brand	24	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)

Note: (1) Differences were calculated as laboratory-measured values minus brand-registered values; therefore, negative values indicate that laboratory-measured concentrations were lower than brand-registered concentrations; (2) The 95% limits of agreement were calculated as the mean difference ± 1.96 standard deviations; (3) NMESs, non-milk extrinsic sugars; SFAs, saturated fatty acids; TFAs, trans fatty acids; LoA, limits of agreement; CI, confidence interval.

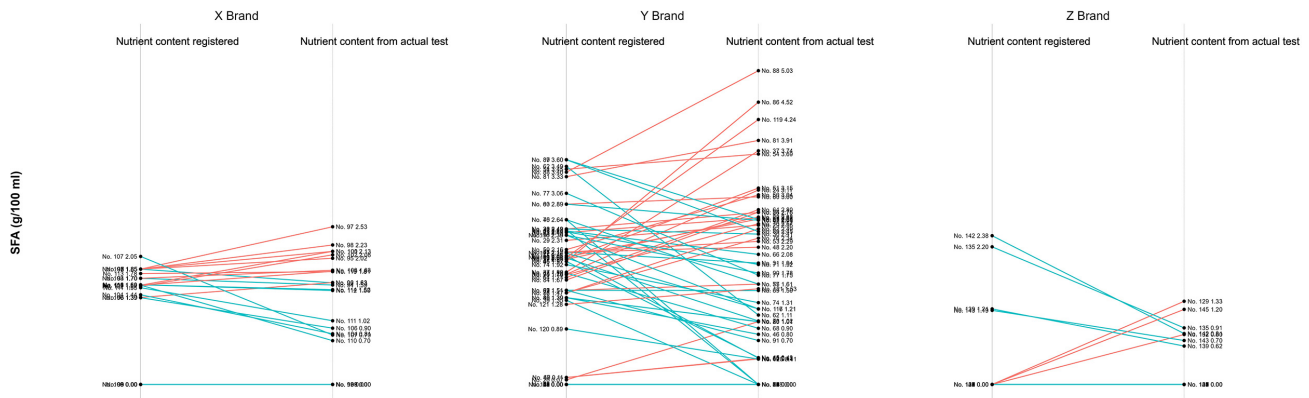


Fig. 8. Differences between $Nutrient_i$ from actual testing and $Nutrient_i$ registered by the brand of various brands of SFAs.

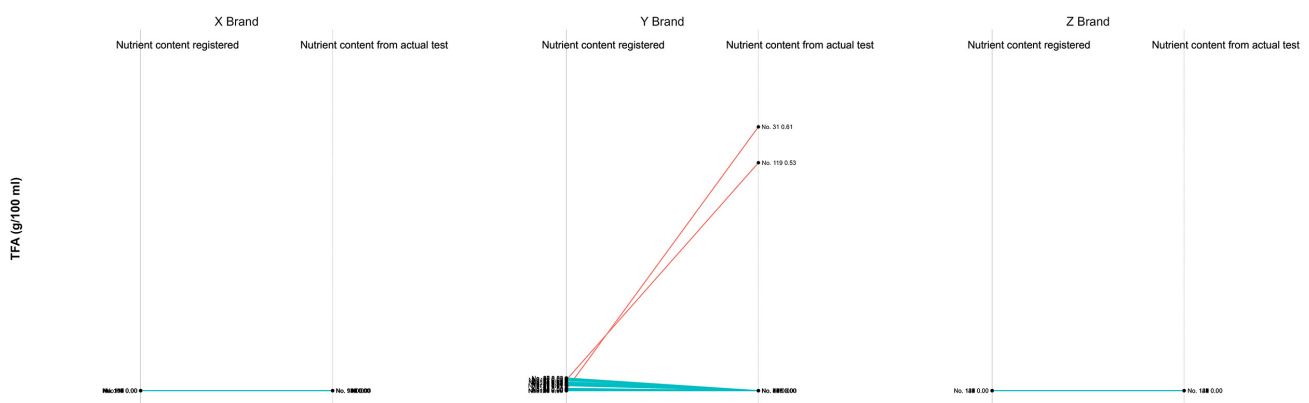


Fig. 9. Differences between $Nutrient_i$ from actual testing and $Nutrient_i$ registered by the brand of various brands of TFAs.

ment coefficients of 100%, 110%, and 125% for Brands X, Z, and Y, respectively, indicating that the tolerance needed to improve grading agreement may differ across retail production settings. Together, these findings provide

preliminary evidence for improving verification and tolerance strategies in the FMB sector.

With regard to the grading results, the overall nutrient composition of the three brands in the present study was

Table 6. The compliance rate between actual grading (AG) and brand-registered grading (BG) with the registered value applied different coefficients.

		Weighted total (n = 160)		X brand (n = 24)		Y brand (n = 74)		Z brand (n = 24)	
		Kappa 95% CI	Compliance rate	Kappa 95% CI	Compliance rate	Kappa 95% CI	Compliance rate	Kappa 95% CI	Compliance rate
Nutrient _i from actual testing	vs	0.688 ^b (0.594, 0.783)	81.88%	0.841 ^c (0.541, 1.000)	95.83%	0.598 ^a (0.465, 0.731)	72.97%	0.611 ^b (0.350, 0.872)	75.00%
Nutrient _i registered by the brand									
Nutrient _i from actual testing	vs	0.754 ^b (0.668, 0.840)	85.53%	0.841 ^c (0.541, 1.000)	95.83%	0.659 ^b (0.530, 0.788)	77.03%	0.739 ^b (0.512, 0.966)	83.33%
Nutrient _i registered by the brand	×								
105%									
Nutrient _i from actual testing	vs	0.789 ^b (0.707, 0.871)	87.42%	0.859 ^c (0.622, 1.000)	95.83%	0.679 ^b (0.552, 0.806)	78.38%	0.875 ^c (0.706, 1.000)	91.67%
Nutrient _i registered by the brand	×								
110%									
Nutrient _i from actual testing	vs	0.790 ^b (0.710, 0.870)	87.42%	0.859 ^c (0.622, 1.000)	95.83%	0.680 ^b (0.553, 0.807)	78.38%	0.875 ^c (0.706, 1.000)	91.67%
Nutrient _i registered by the brand	×								
115%									
Nutrient _i from actual testing	vs	0.812 ^c (0.736, 0.888)	88.68%	0.859 ^c (0.622, 1.000)	95.83%	0.720 ^b (0.597, 0.844)	81.08%	0.875 ^c (0.706, 1.000)	91.67%
Nutrient _i registered by the brand	×								
120%									
Nutrient _i from actual testing	vs	0.855 ^c (0.784, 0.926)	91.19%	0.859 ^c (0.622, 1.000)	95.83%	0.800 ^c (0.688, 0.912)	86.49%	0.875 ^c (0.706, 1.000)	91.67%
Nutrient _i registered by the brand	×								
125%									
Nutrient _i from actual testing	vs	0.865 ^c (0.796, 0.934)	91.82%	0.859 ^c (0.622, 1.000)	95.83%	0.820 ^c (0.712, 0.928)	87.84%	0.875 ^c (0.706, 1.000)	91.67%
Nutrient _i registered by the brand	×								
130%									

Note: (1) ^a moderate agreement, ^b high agreement, ^c almost perfect agreement; (2) All Kappa statistics are statistically significant with $p < 0.001$; (3) Brand-registered values were multiplied by coefficients from 105% to 130% in 5% increments and compared with laboratory-measured values to recalculate grading consistency rates; (4) AG, actual grading (derived from laboratory-measured nutrient values); BG, brand-registered grading (derived from brand-registered nutrient values); (5) The actual sample size is $n = 122$. The “Weighted total ($n = 160$)” is the effective weighted sample size used in the weighted Kappa analysis, not the actual number of collected samples. Compliance rate was defined as the proportion of samples in which AG was consistent with or better than BG, reflecting the unidirectional tolerance framework for restricted nutrients. It was calculated as $\text{Compliance rate (\%)} = \frac{N_{\text{AG-BG matches}}}{N_{\text{total}}} \times 100\%$.

Table 7. Post hoc pairwise comparisons following the Kruskal-Wallis tests.

Grouping variable	Nutrient	Group 1	Group 2	Difference in mean ranks	Standard error	Z	Unadjusted <i>p</i> value	Bonferroni-adjusted <i>p</i> value
Brand	NMESs	X	Y	−22.011	8.271	−2.661	0.008	0.023
		X	Z	−23.208	10.164	−2.283	0.022	0.067
		Y	Z	−1.197	8.271	−0.145	0.885	1.000
	SFAs	X	Y	−0.766	8.096	−0.095	0.925	1.000
		X	Z	33.708	9.949	3.388	0.001	0.002
		Y	Z	34.474	8.096	4.258	<0.001	<0.001
Nutri-Choice classification	NMESs	A	B	−26.956	15.112	−1.784	0.074	0.447
		A	C	−86.081	15.704	−5.481	<0.001	<0.001
		A	D	−61.143	15.840	−3.860	<0.001	0.001
		B	C	−59.125	7.858	−7.525	<0.001	<0.001
		B	D	−34.187	8.126	−4.207	<0.001	<0.001
		D	C	24.938	9.180	2.717	0.007	0.040
	SFAs	A	B	−27.193	14.792	−1.838	0.066	0.396
		A	C	−8.774	15.371	−0.571	0.568	1.000
		A	D	−59.107	15.504	−3.812	<0.001	0.001
		B	C	18.419	7.691	2.395	0.017	0.100
		B	D	−31.914	7.954	−4.013	<0.001	<0.001
		D	C	−50.333	8.985	−5.602	<0.001	<0.001
Declared sweetness	NMESs	No extra sugar	Quarter sugar	24.531	9.180	2.672	0.008	0.075
		No extra sugar	Half sugar	25.104	10.493	2.392	0.017	0.167
		No extra sugar	Less sugar	42.045	9.302	4.520	<0.001	<0.001
		No extra sugar	Full sugar	61.888	9.574	6.464	<0.001	<0.001
		Quarter sugar	Half sugar	0.573	11.364	0.050	0.960	1.000
		Quarter sugar	Less sugar	17.514	10.274	1.705	0.088	0.882
		Quarter sugar	Full sugar	37.357	10.521	3.551	<0.001	0.004
		Half sugar	Less sugar	16.942	11.462	1.478	0.139	1.000
		Half sugar	Full sugar	36.784	11.684	3.148	0.002	0.016
		Less sugar	Full sugar	19.843	10.627	1.867	0.062	0.619
Drink types	NMESs	Pure tea	Fruit Tea & Fruit juice	−35.310	12.171	−2.901	0.004	0.011
		Pure tea	Milk coffee & Milk tea	2.325	11.358	0.205	0.838	1.000
		Fruit Tea & Fruit juice	Milk coffee & Milk tea	37.635	7.193	5.233	<0.001	<0.001
	SFAs	Pure tea	Fruit Tea & Fruit juice	−3.229	11.913	−0.271	0.786	1.000
		Pure tea	Milk coffee & Milk tea	−60.316	11.118	−5.425	<0.001	<0.001
		Fruit Tea & Fruit juice	Milk coffee & Milk tea	−57.087	7.040	−8.109	<0.001	<0.001

Table 8. Distribution of the RE of samples (N = 122) (N (%)).

Classification	Subunit	NMESs				SFAs				TFAs				N
		Median [Q1, Q3]	Rank-sum test		Reach the claimed (RE ≤0.2)	Median [Q1, Q3]	Rank-sum test		Reach the claimed (RE ≤0.2)	Median [Q1, Q3]	Rank-sum test		Reach the claimed (RE ≤0.2)	
			H value	p value			H value	p value			H value	p value		
Brand	X brand	−0.34 [−1.00, −0.23]	8.67	0.013	24 (100.00)	0.00 [−0.11, 0.15]	1.16	0.599	20 (83.33)	0.00 [−0.00, 0.00] ^c	47.35	<0.001	24 (100.00)	24
	Y brand	−0.23 [−0.54, 0.04] ^a			63 (85.14)	0.00 [−0.27, 0.18]			57 (77.03)	−1.00 [−1.00, 0.00]			72 (97.30)	74
	Z brand	−0.07 [−0.39, 0.05] ^a			22 (91.67)	0.00 [−0.00, 0.00]			24 (100.00)	0.00 [−0.00, 0.00] ^c			24 (100.00)	24
Drink types	Pure tea	−0.18 [−0.37, 0.04]	14.75	0.002	11 (100.00)	0.00 [−0.00, 0.00]	1.23	0.745	11 (100.00)	0.00 [−0.00, 0.00] ^d	39.46	<0.001	11 (100.00)	11
	Fruit Tea	−0.06 [−0.31, 0.16] ^b			23 (79.31)	0.00 [−0.00, 0.00]			26 (89.66)	0.00 [−0.00, 0.00] ^d			29 (100.00)	29
	Fruit juice	−0.07 [−0.41, 0.15]			5 (83.33)	0.00 [−0.00, 0.00]			6 (100.00)	0.00 [−0.00, 0.00] ^d			6 (100.00)	6
	Milk tea/milk coffee	−0.32 [−1.00, −0.12]			70 (92.11)	−0.02 [−0.42, 0.19]			58 (76.32)	−1.00 [−1.00, 0.00]			74 (97.37)	76
Weighted total		−0.25 [−0.66, −0.04]			137 (85.63)	0.00 [−0.13, 0.09]			127 (79.38)	0.00 [−1.00, 0.00]			158 (98.75)	160

RE, Relative error. Values are presented as median [Q1, Q3]. Different superscript letters within the same column indicate statistically significant differences among groups ($p < 0.05$). Groups sharing at least one common superscript letter are not significantly different.

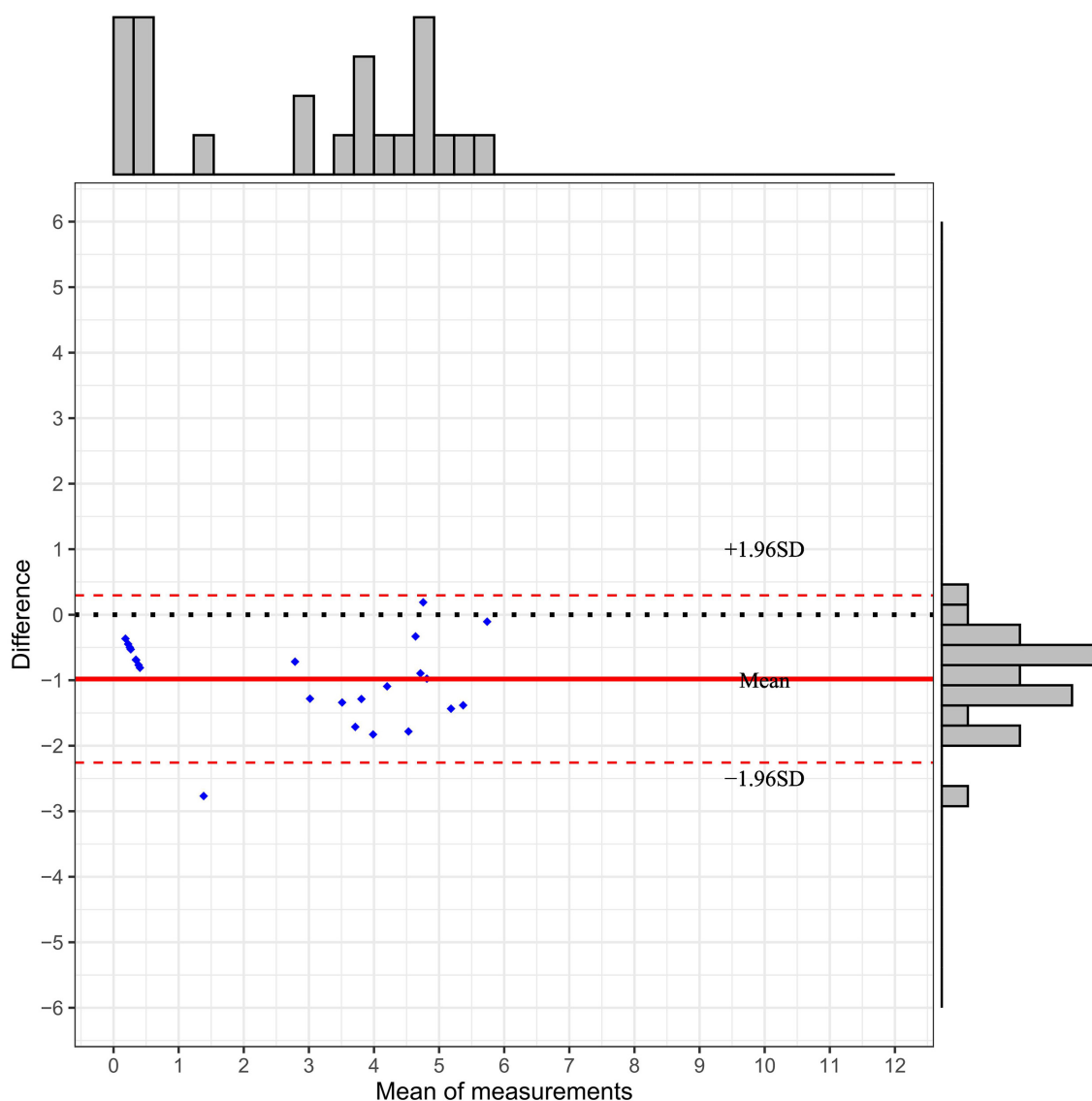


Fig. 10. Bland-Altman plot of brand-registered and laboratory-measured NMESs values for Brand X. Note: Each point represents one beverage sample. The x-axis shows the mean of the paired values, and the y-axis shows the difference calculated as laboratory-measured minus brand-registered. The solid red line represents the mean difference, the dashed red lines represent the 95% limits of agreement, and the dotted horizontal line represents zero difference.

compared with the findings reported by Shi et al. [8] for retail handcrafted milk tea products in Shanghai. In the present samples, the median NMESs and TFAs concentrations were 3.63 g/100 mL and 0.00 g/100 mL, respectively, whereas Shi et al. [8] reported median total sugar and trans-fat levels of 6.50 g/100 g and 0.721 g/100 g, respectively. From Grades A to C, NMESs concentrations tended to increase. However, most Grade D beverages were assigned to Grade D because their SFA content exceeded 2.5 g/100 mL. This finding is consistent with the current nutritional profile of FMBs in China. Most beverages had NMESs concentrations below 11.5 g/100 mL, while some beverages contained relatively high SFAs levels, especially those with milk foam, cream, or dairy-based ingredients. TFAs were

also generally absent or detected only at very low levels in the sampled beverages. Moreover, although Brand Y involved the highest proportion of manual preparation among the three brands, its median NMESs level was not the highest; this suggests that NMESs content is mainly determined by product formulation and standardized recipes rather than by the extent of manual intervention.

Nutrition-labeling regulations in many countries have set scientifically based tolerance ranges for declared nutrient values, commonly limiting restricted components (e.g., fat, SFAs, TFAs, NMESs, sodium) to $\leq 120\%$ of the labeled value to allow for analytical variability and production heterogeneity [9,25,26]. However, evidence supporting error coefficients for nutrient content in FMBs remains

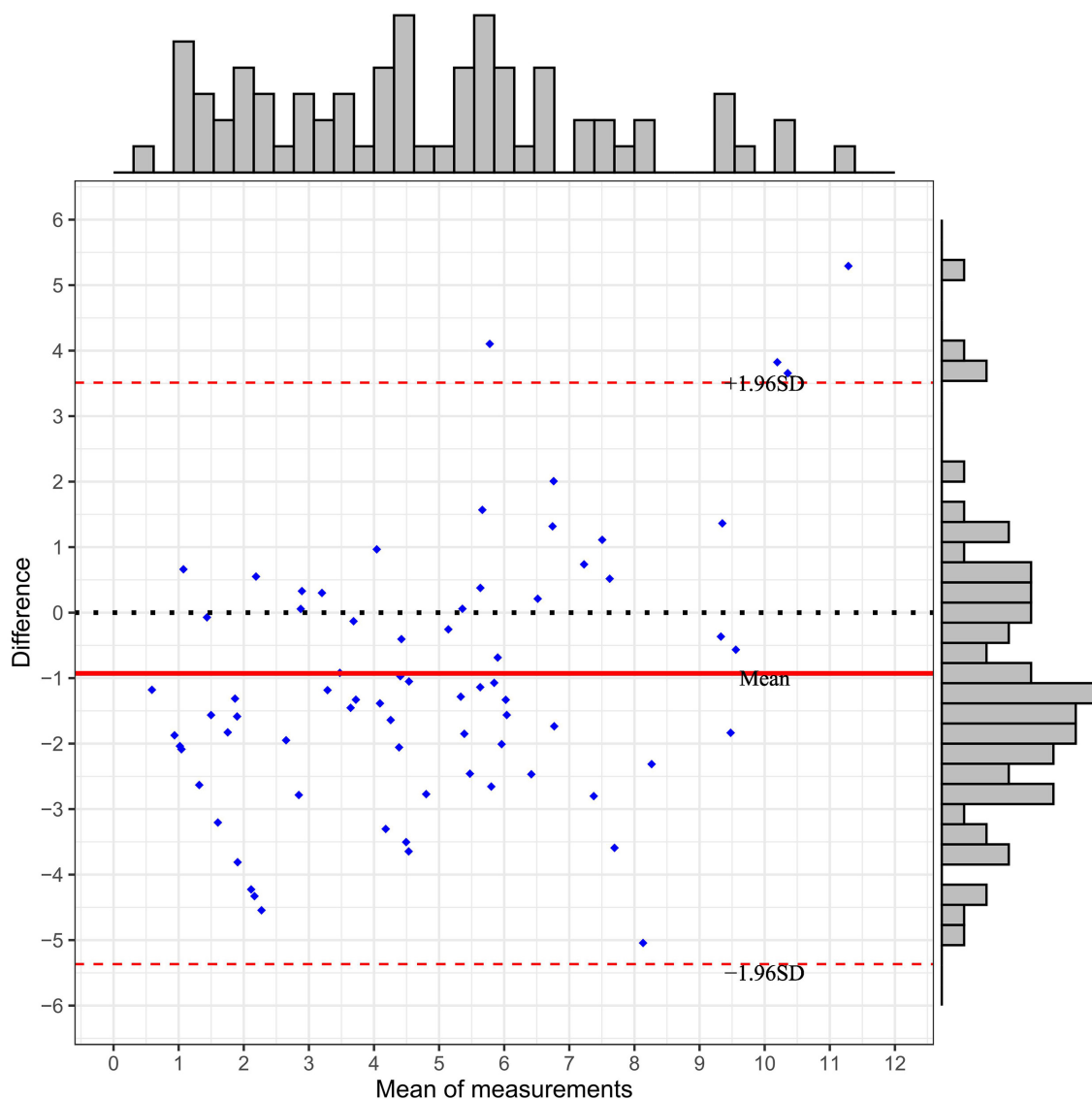


Fig. 11. Bland-Altman plot of brand-registered and laboratory-measured NMESs values for Brand Y. Note: Each point represents one beverage sample. The x-axis shows the mean of the paired values, and the y-axis shows the difference calculated as laboratory-measured minus brand-registered. The solid red line represents the mean difference, the dashed red lines represent the 95% limits of agreement, and the dotted horizontal line represents zero difference.

limited. To reduce possible labeling inaccuracies in FMBs, Singapore's Nutri-Grade implementation framework uses a precautionary strategy by multiplying pre-ice formulations of customized beverages (e.g., bubble tea, soymilk). This strategy is designed to account for nutrient-content variability arising from artisanal beverage preparation while preserving regulatory compliance. Nevertheless, systematic evidence regarding the suitability of the 120% tolerance threshold for contemporary FMBs is still insufficient. In this context, the present findings offer empirical support for improving nutrition-labeling accuracy in freshly made beverages, a critical yet still underregulated sector.

Beyond nutrient reformulation, the performance of a FOPL system also relies on consumer engagement and trust. Trust has been recognized as a key mediator between

consumer attitudes toward FOPL and their behavioral intentions [28]. Labels that are perceived as accurate and reliable are more likely to influence purchasing decisions and promote healthier choices [29]. This consideration is especially relevant for FMBs, because these products are prepared on-site and cannot be checked by consumers before purchase. Therefore, the credibility of Nutri-Choice largely depends on whether the displayed grade accurately represents the beverage actually consumed. Preserving this trust requires continuous labeling accuracy, robust verification mechanisms, and, when needed, production-specific adjustment strategies. In this regard, the present study provides empirical evidence supporting verification-based monitoring and production-specific tolerance strategies for FMB nutrition labeling.

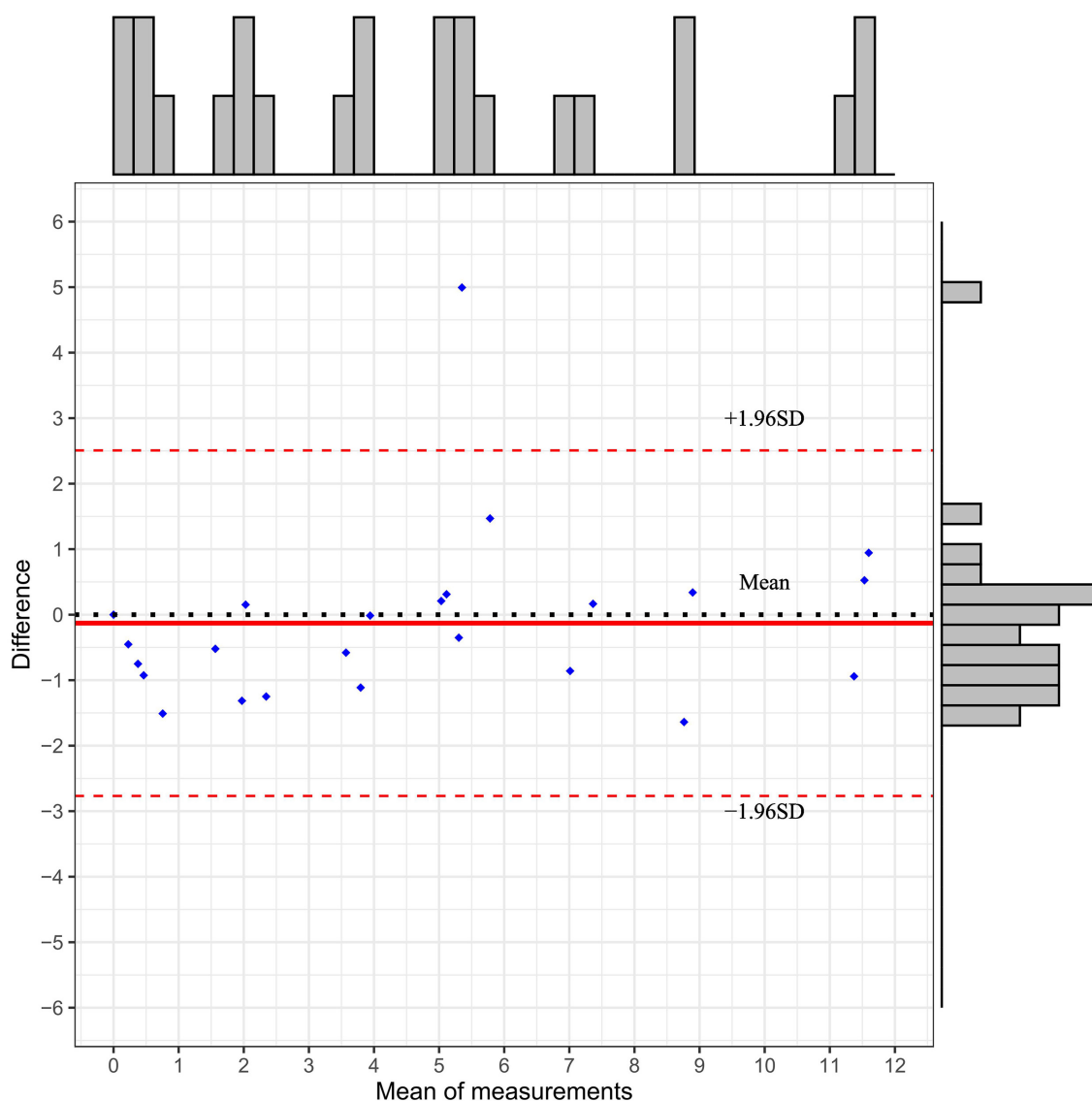


Fig. 12. Bland-Altman plot of brand-registered and laboratory-measured NMESs values for Brand Z. Note: Each point represents one beverage sample. The x-axis shows the mean of the paired values, and the y-axis shows the difference calculated as laboratory-measured minus brand-registered. The solid red line represents the mean difference, the dashed red lines represent the 95% limits of agreement, and the dotted horizontal line represents zero difference.

Strengths and Limitations

This study has several strengths. The three selected brands had different operational characteristics and represented the main FMB product types, allowing descriptive comparisons across real-world retail contexts. By auditing beverages purchased under routine retail conditions instead of depending on manufacturer-declared or laboratory-prepared values, this study captured the operational variability inherent in point-of-sale preparation, a factor that has been largely neglected in earlier nutrient profiling and FOPL studies. In contrast to most previous studies focused on pre-packaged products, this study evaluated whether manufacturer-submitted nutrient grades remain accurate under actual retail conditions. In addition, while

existing tolerance frameworks use uniform margins across production types, the brand-stratified analysis offers a practical basis for assessing whether differentiated tolerances may improve grading agreement, which has not previously been documented in the FMB sector.

Several limitations should be considered. First, the geographic focus on Shanghai, the voluntary design of the pilot, and the inclusion of only three brands restrict the generalizability of the findings. Second, although the sample size was acceptable for exploratory analysis, it remains limited for defining definitive tolerance thresholds, and the 9% margin of error together with unequal sampling fractions may have introduced imprecision; although brand-level sampling weights were used in the weighted Kappa analysis, residual bias cannot be excluded. Third, each pro-

duction profile was represented by only one brand; consequently, brand effects and production-method effects were confounded. The observed differences may be related to product formulation, product mix, ingredient sourcing, employee training, store practices, production methods, or several of these factors together. In view of these limitations, the proposed adjustment coefficients should be regarded as exploratory and require validation in broader populations before generalization. Future multi-center studies including more brands, regions, and seasonal variation are needed to validate and refine these coefficients.

5. Conclusion

Nutri-Choice provides a practical framework for applying graded nutrition labeling to FMBs, but its credibility depends on routine market sampling, laboratory verification, and evidence-based tolerance strategies. The variation observed in grading agreement among brands suggests that differentiated tolerance strategies should be evaluated further. Future multicentre studies should involve multiple brands within each preparation modality and include broader regional, seasonal, and store-level coverage to distinguish brand-level characteristics from preparation-related variability. Further research on consumer understanding and use of the Nutri-Choice label is also required to support future nutrition-labeling policies for FMBs.

Abbreviations

FMBs, Freshly made beverages; O2O, Online-to-offline; SSBs, Sugar-sweetened beverages; TFAs, Trans fatty acids; NMESs, Non-milk-extrinsic sugars; SFAs, Saturated fatty acids; NSSs, Non-sugar sweeteners; SOPs, Standard operating procedures; FOPL, Front-of-package labeling; SCDC, Shanghai Municipal Center for Disease Control and Prevention; CNAS, China National Accreditation Service for Conformity Assessment; SKUs, Stock-keeping units; CMA, China Metrology Accreditation; HPLC, High-performance liquid chromatography; RE, Relative error; AG, Actual grading; BG, Brand-registered grading; LoA, Limits of agreement.

Availability of Data and Materials

The data that support the findings of this study are available from the corresponding authors on reasonable request.

Author Contributions

ZSun, ZShi, and YW contributed to the conception, designed of the work, developed the analysis and drafted the manuscript. ZW and JZ led the conceptual framework. ZShi, WM, LS, SM, QS, WL collected the data. ZSun, WM, LS, SM, QS, WL offered support of cleaning data. ZW, JZ critically revised the manuscript for important intellectual content. XL performed re-analysis and verification

of the key data, and assisted in improving the text and figures during the revision. All authors discussed questions of accuracy or integrity of the work, ensured differences were on appropriately investigated and resolved, and agreed on the version to be published. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

This research has been conducted in Shanghai Nutrition and Health Guidance pilot project. We are grateful to all the participants and staff who have been involved in the programs.

Funding

Z.S. was supported by the National Natural Science Foundation of China (82607224), Shanghai Sailing Program (23YF1437000); J.Z. was supported by the National Natural Science Foundation of China (72674195), Yum China Dietary Health Foundation (CNS-YUM2024-122), Shanghai Oriental Talents (Excellent Academic Leader Program) (BJWS2024048).

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration of AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work the authors used KIMI in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Appendix

See Tables 7,8; Figs. 7,8,9,10,11,12.

References

- [1] China Chain Store & Franchise Association. The 2023 New Tea Research Report. 2023. Available at: <http://www.ccfa.org.cn/portal/cn/xiangxi.jsp?id=444980&type=33> (In Chinese) (Accessed: 5 July 2026).
- [2] Zheng M, Zhang J, He H, Xu Y, Ma G. SWOT analysis on tax on sugary beverages in China. *Food and Nutrition in China*. 2017; 23: 38–41. <https://doi.org/10.3969/j.issn.1006-9577.2017.05.009> (In Chinese)
- [3] Pan F, Luan D, Zhang T, Mao W, Liang D, Liu A, et al. Assessment of sugar-sweetened beverages consumption and free sugar intake among urban residents aged 3 and above in China. *Chinese Journal of Food Hygiene*. 2022; 34: 126–130. <https://doi.org/10.13590/j.cjfh.2022.01.024> (In Chinese)

- [4] Malik VS, Hu FB. The role of sugar-sweetened beverages in the global epidemics of obesity and chronic diseases. *Nature Reviews. Endocrinology*. 2022; 18: 205–218. <https://doi.org/10.1038/s41574-021-00627-6>
- [5] Huang Y, Chen Z, Chen B, Li J, Yuan X, Li J, et al. Dietary sugar consumption and health: umbrella review. *BMJ (Clinical Research Ed.)*. 2023; 381: e071609. <https://doi.org/10.1136/bmj-2022-071609>
- [6] Chinese Nutrition Society. *Dietary Guidelines for Chinese Residents* (2021). Chinese Nutrition Society: Beijing, China. 2021. (In Chinese)
- [7] National Health Commission of the People's Republic of China. *Nutrition and Chronic Disease Status of Chinese Residents* (2020). People's Medical Publishing House: Beijing, China. 2020. (In Chinese)
- [8] Shi Z, Sun Z, Song Q, Qu M, Wang Z, Zang J. Nutrient content of 122 kinds of retail handcrafted milk tea products in Shanghai. *Journal of Environmental and Occupational Medicine*. 2023; 40: 756–760, 768. <https://doi.org/10.11836/JEOM22505> (In Chinese)
- [9] National Health Commission of the People's Republic of China, State Administration for Market Regulation. *National Food Safety Standard—General Principles for Nutrition Labelling of Prepackaged Foods: GB 28050-2025*. Standards Press of China: Beijing, China. 2025. (In Chinese)
- [10] Shanghai Municipal Center for Disease Control & Prevention. *How to choose the Nutri-Choice Labelling*. 2024. Available at: <https://mp.weixin.qq.com/s/ZC6nld9kK2sSrjr2fCgYzw> (In Chinese) (Accessed: 5 July 2026).
- [11] Health Promotion Board. *A Handbook on Nutrition Labelling* (Singapore). Health Promotion Board: Singapore, Singapore. 2022.
- [12] Sullivan F. *New Style Tea Drink Supply Chain Mid-Term and Long-Term Development Study: China's New Style Tea Drink Supply Chain White Paper*. 2022. Available at: <https://www.frostchina.com/en/content/insight/detail/694e28be4a7a7390dece6061> (In Chinese) (Accessed: 5 July 2026).
- [13] Liu Y, Shi M, Pan F, Li G, Luan D, Liu A, et al. A study on free sugar content in commercially available beverages in China. *Chinese Journal of Food Hygiene*. 2021; 33: 93–96. <https://doi.org/10.13590/j.cjfh.2021.01.018> (In Chinese)
- [14] Li Y, Wang H, Zhang P, Popkin BM, Coyle DH, Ding J, et al. Nutritional Quality of Pre-Packaged Foods in China under Various Nutrient Profile Models. *Nutrients*. 2022; 14: 2700. <https://doi.org/10.3390/nu14132700>
- [15] National Standard of the People's Republic of China. *General Standard for Beverage: GB/T 10789-2015*. Standards Press of China: Beijing, China. 2015. (In Chinese)
- [16] National Standard of the People's Republic of China. *Tea Beverages: GB/T 21733-2008*. Standards Press of China: Beijing, China. 2008. (In Chinese)
- [17] National Standard of the People's Republic of China. *Fruit and Vegetable Juice Beverages: GB/T 31121-2014*. Standards Press of China: Beijing, China. 2014. (In Chinese)
- [18] National Standard of the People's Republic of China. *Coffee Beverages: GB/T 30767-2014*. Standards Press of China: Beijing, China. 2014. (In Chinese)
- [19] National Health Commission of the People's Republic of China. *National Food Safety Standard—Determination of Fructose, Glucose, Sucrose, Maltose, and Lactose in Foods: GB 5009.8-2023*. Standards Press of China: Beijing, China. 2023. (In Chinese)
- [20] National Health and Family Planning Commission of the People's Republic of China. *National Food Safety Standard—Determination of Fatty Acids in Foods: GB 5009.168-2016*. Standards Press of China: Beijing, China. 2016. (In Chinese)
- [21] National Health and Family Planning Commission of the People's Republic of China. *National Food Safety Standard—Determination of Trans Fatty Acids in Foods: GB 5009.257-2016*. Standards Press of China: Beijing, China. 2016. (In Chinese)
- [22] Chinese Nutrition Society. *Specification for Nutri-Choice Labelling of Freshly Made Beverages: T/CNSS 045-2025*. 2025. Available at: <https://www.nutritionsciences.org.cn/nutrition/scientific/paperdetail/452?id=53&key=guide> (In Chinese) (Accessed: 5 July 2026).
- [23] Health Promotion Board. *Nutri-Grade Mark for Nutri-Grade Beverages: Usage Guide for Graphic Applications. Version 2.1*. Health Promotion Board: Singapore, Singapore. 2023.
- [24] Macdonald IA. Free sugars. *The Proceedings of the Nutrition Society*. 2020; 79: 56–60. <https://doi.org/10.1017/S0029665119001046>
- [25] Centre for Food Safety, Food and Environmental Hygiene Department. *Technical Guidance Notes on Nutrition Labelling and Nutrition Claims*. Hong Kong; August 2016.
- [26] 21 CFR 101.9. *Nutrition labeling of food*. 2025. Available at: <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-B/part-101/subpart-A/section-101.9> (Accessed: 5 July 2026).
- [27] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33: 159–174. <https://doi.org/10.2307/2529310>
- [28] Mazzù MF, Baccelloni A, Romani S, Andria A. The role of trust and algorithms in consumers' front-of-pack labels acceptance: a cross-country investigation. *European Journal of Marketing*. 2022; 56: 3107–3137. <https://doi.org/10.1108/ejm-10-2021-0764>
- [29] Fuller K, Sharib JR, Fan B, Pomeranz JL, Mozaffarian D, Cash SB. U.S. Consumer Preferences for FDA Healthy vs. Generic Healthy Food Labels: The Influence of Trust. *Food Quality and Preference*. 2026; 144: 105898. <https://doi.org/10.1016/j.foodqual.2026.105898>