Electronic Supplementary Material (ESI) for Green Chemistry.
This journal is © The Royal Society of Chemistry 2022

# Supplementary material

# Original end-to-end smart diagnosis framework of systematic critical quality attributes benchmarking FDA standard of phytomedicine by biosensor and multi-information fusion coupled with AI algorithm

This Supplementary material contains 11 tables and one figure.

## 1. Materials and methods

### 1.1 Instruments and materials

The 30 batches of alcohol-precipitated intermediates, corresponding 30 batches of water-precipitated intermediates, and corresponding 30 batches of products of Xiaoer Xiaoji Zhike oral liquid were provided by Lunan Pharmaceutical Group Co., Ltd., all of which were real-world materials. The specific lot number was shown as follows: 130121164, 130121165, 130121166, 130121167, 130121168, 130121169, 130121170, 130121171, 130121172, 130121173, 130121174, 130121175, 130121177, 130121178, 130121179, 130121180, 130121181, 130121182, 130121184, 130121185, 130121187, 130121188, 130121189, 130121190, 130121192, 130121224, 130121225, 130121226, 130121227, 130121228.

The diagnosis of additional 30 batches of samples to verify reliability.（Lot number: 204200112, 204200122, 204200102, 204200062, 204200072, 204200182, 204200092, 204200152，304200383, 304200483, 304200403, 304200353, 304200393, 304200423, 304200563, 304200553, 304200473, 204200162, 304200413, 304200533, 304200543, 304200503, 204200132, 304200433, 304200373, 304200523, 304200493, 304200463, 204200142, 304200513）

### 1.2 Biological CQAs digitization and smart diagnosis for 30 batches of Xiao'er Xiaoji Zhike oral liquid in end-to-end real world by MIF-HEMT biosensor integrated UPLC-MS/MS

**Table S1.** The gradient elution table of UPLC-MS/MS for Biological CQAs digitization

| No. | Retention (min) | Flow (mL·min$^{-1}$) | %A | %B |
|---|---|---|---|---|
| 1 | 0 | 0.3 | 99 | 1 |
| 2 | 0 | 0.3 | 99 | 1 |
| 3 | 1 | 0.3 | 99 | 1 |
| 4 | 6 | 0.3 | 90 | 10 |
| 5 | 15 | 0.3 | 82 | 18 |
| 6 | 25 | 0.3 | 70 | 30 |
| 7 | 28 | 0.3 | 50 | 50 |
| 8 | 31 | 0.3 | 20 | 80 |
| 9 | 32 | 0.3 | 0 | 100 |
| 10 | 35 | 0.3 | 0 | 100 |
| 11 | 37 | 0.3 | 99 | 1 |
| 12 | 40 | 0.3 | 99 | 1 |

Chromatographic column electrospray ion source (ESI) mode of positive and negative ions. The Fourier high-resolution scanning range was m/z from 100 to 1200, and the primary resolution was 30000. The secondary mass spectrometry adopted data dependent scanning, and selected the three ions with the highest primary abundance for CID secondary fragmentation. When the secondary fragment information was incomplete, the ion list scanning method was used to improve the acquisition efficiency of secondary mass spectrometry information. Flow rate was set as 0.30 mL·min-1 and injection volume as 3 μL. Using 0.1% formic acid water and acetonitrile as mobile phases. The sheath gas flow rate was set to 40 ARB, the auxiliary gas flow rate to 20 ARB. The capillary voltage was set as -35 V, spray voltage as 3 kV and the tube lens voltage as -110 V. The capillary temperature was 350 centigrade. The activation energy unit was set to 0.25 Q and the activation time was 30 ms. The normalized collision energy was 35%. Finally, Xcalibur 2.1 workstation was used for data processing coupled with molecular prediction module. The parameters were set as, C [0-20], H [0-30], O [0-15], n [0-3], s [0-1], number of rings and unsaturated bonds [0-15], and the mass accuracy error was within 10.

### 1.3 Chemical CQAs digitization and smart diagnosis of 30 batches of end-to-end real-world Xiaoer Xiaoji Zhike oral liquid from three pharmaceutical units

According to the Pharmacopoeia of the People's Republic of China (2020) and previous basis, chemical CQAs and physical CQAs are essential to product quality in ETE-SDF. Based on the aforementioned biological CQAs, chemical CQAs were selected and further monitored. 30 batches of

Xiao'er Xiaoji Zhike oral liquid from three pharmaceutical units in end-to-end real world was implemented by UPLC. The 200 μL alcohol-precipitated intermediate, water-precipitated intermediate, and product of Xiaoer Xiaoji Zhike oral liquid were precisely measured and put in a flask with a volume of 2 mL, respectively. Then water was added to the mark. And the samples were shaken well and filtered through a 0.22 μm microporous membrane for measurement.

Chemical CQAs digitization and 30 batches of end-to-end real-world Xiaoer Xiaoji Zhike oral liquid from three pharmaceutical units were implemented by UPLC. The linear gradient elution program was shown in **Table S2**.

**Table S2.** The linear gradient elution program of samples from three pharmaceutical units

| Sample | Retention (min) | A% | B% | Flow ($mL \cdot min^{-1}$) | Temperature (℃) |
|---|---|---|---|---|---|
| Water-precipitation intermediates and products | 0 | 100 | 0 | 0.3 | 35 |
| | 3 | 100 | 0 | 0.3 | 35 |
| | 6 | 99 | 1 | 0.3 | 35 |
| | 15 | 88 | 12 | 0.3 | 35 |
| | 48 | 79 | 21 | 0.3 | 35 |
| | 59 | 70 | 30 | 0.3 | 35 |
| | 66 | 5 | 95 | 0.3 | 35 |
| Alcohol-precipitation intermediates | 0 | 100 | 0 | 0.3 | 35 |
| | 3 | 100 | 0 | 0.3 | 35 |
| | 6 | 99 | 1 | 0.3 | 35 |
| | 15 | 88 | 12 | 0.3 | 35 |
| | 53 | 79 | 21 | 0.3 | 35 |
| | 64 | 70 | 30 | 0.3 | 35 |
| | 70 | 5 | 95 | 0.3 | 35 |

Method validation for standard analysis of alcohol-precipitation intermediates is as follows.

(1) Linearity

Linear regression analysis of each of the four compounds (Forsythiaside E, Neoeriocitrin, Hesperidin, Neohesperidin) was performed in triplicate using six different concentrations.

The line for each compound was plotted using linear regression of the peak area vs concentration. $y = ax + b$, x indicated the concentrations of the marker compounds ($\mu g \cdot mL^{-1}$), y and $R^2$ were the peak area and coefficient of correlation of the equation, respectively. The $R^2$ was used to determine the linearity. All the marker compounds showed linearity ($R^2 > 0.999$) in the results shown in **Table S5**.

(2) Limit of detection (LOD) and limit of quantification (LOQ)

The LOD and LOQ were determined at a signal-to-noise ratio (S/N) of 3 and 10, respectively as the lowest concentrations of the analyte.

(3) Precision

The Precision was determined from six analyses on the same day.

(4) Repeatability

The repeatability of the developed method was estimated by sampling six times a day.

(5) Stability

Furthermore, the stability was analyzed by injecting three aliquots of a sample solution during six-time points, 0h, 3h, 6h, 12h 18h, and 24h. The RSD was considered a measure of precision.

(6) Recovery

An appropriate amount of the Xiao'er Xiaoji Zhike oral liquid was divided into one portion as the control group, and another three portions that were spiked with marker standards at three concentration levels (80%, 100%, and 120%). The filtrates were assayed using UPLC to determine the recoveries, which were calculated using the following equation:

Recovery (%) = (total amount detected - amount original)/amount spiked × 100

Similarly, the results of the methodological validation of water precipitation intermediates and products were shown in **Table S5**.

In addition, the methodological verification of chromatographic fingerprint was as follows.

(1) Precision

After repeated determination of six chromatograms of the test article, the similarity of alcohol precipitation intermediates, water precipitation intermediates, and finished products was calculated using the software of *the similarity evaluation system of chromatographic fingerprint of traditional Chinese medicine (2012 version).*

(2) Repeatability

Six test solutions were used respectively, and the similarity of alcohol precipitation intermediates, water precipitation intermediates, and finished products was calculated by using the software of *the similarity evaluation system of chromatographic fingerprint of traditional Chinese medicine (2012 Edition).*

(3) Stability

One test solution was injected at 0, 2, 8, 16, 24, and 48 h respectively. The similarity of alcohol precipitation intermediates, water precipitation intermediates, and finished products was calculated using the software of *the similarity evaluation system of chromatographic fingerprints of traditional Chinese medicine (2012 Edition)*.

Similarly, the results of the methodological validation of three units of fingerprint analysis were shown in **Table S6**.

**1.4 Physical CQAs digitization and smart diagnosis for physical CQAs of 30 batches of end-to-end real-world Xiaoer Xiaoji Zhike oral liquid from three pharmaceutical units**

The product of Xiao'er Xiaoji Zhike oral liquid was measured repeatedly for three times in a cuvette, and the average L*, a*, b* values and Eab value of the 3 times were calculated and adopted.

$$E_{ab} = (L^2 + a^2 + b^2)^{1/2} \quad (4)$$

Taste is a vital quality attribute of the oral preparation. It reflects the adaptability of drug quality, which plays a decisive role in patient compliance, especially for children, and directly affects its market sales. The electronic tongue is a modern qualitative quantitative analysis detection instrument for taste measurement, which mainly composed of the interactive sensitive sensor array, signal acquisition circuit, and pattern recognition-based data processing methods.

As a taste bionic technology, the C-tongue electronic tongue was introduced and blazed new trails for the taste characterization of 30 batches of real-world Xiaoer Xiaoji Zhike oral liquid samples. This type of electronic tongue is mainly composed of a stable sensor array with seven metal electrodes. The original idea of a combined pulse relaxation spectrum is realized through voltammetry electrochemical pulse technology excitation. Through interactive induction analysis technology, the overall information of the measured object is obtained.

More specifically, (1) Working electrode composition: platinum electrode, gold electrode, palladium electrode, titanium electrode, tungsten electrode, silver electrode.

(2) Auxiliary electrode: platinum electrode.

(3) Reference electrode: platinum electrode and Ag/AgCl electrode.

(4) Signal acquisition: high-frequency relaxation pulse signal, from +1 v to -1 v, 0.2 v/time.

(5) Frequency of the pulse signal: 1 Hz, 10 Hz, 100 Hz.

(6) Time interval of the pulse signal: 0.001 s.

(7) Data magnification factor: up to $10^6$.

(8) Signal excitation acquisition system: sampling rate $\geqslant$ 1 kb·$s^{-1}$.

(9) Scanning sensitivity: $10^{-6}$ M.

(10) Hardware requirements: the sensor has stable performance, good reproducibility, long service life, rich detection information, 2 - 3 min cleaning time of sensors, which lasts 2.6 s at 0 V, 2.6 s at 1.2 v, 2.6 s at -1.2 v, and improves the stability of detection.

(11) Signal description: the signal collected is the overall response of the sample, rather than the results of the concentration of a specific component.

Specific steps of the electronic tongue test for Xiaoer Xiaoji Zhike oral liquid are as follow.

Step 1. Electrode preheating. Taking 20 mL of deionized water with a sensitivity magnification of 100 times, the eight electrodes of the electronic tongue were preheated for 15 min.

Step 2. Sample test. The 2 mL product of Xiaoer Xiaoji Zhike oral liquid was measured precisely to a 20 mL volumetric flask for constant volume. The taste of 30 batches of the samples from three manufacturing units was measured by the C Tongue series electronic tongue with voltage acquisition mode at room temperature. Three samples were prepared for each batch. The data acquisition resolution was set as 16 bits and the acquisition voltage ranged from -10 v to 10 v. In particular, care was taken to guarantee that the electrodes did not contact the vessel during the measurement.

Step 3. Electrode cleaning. 20mL deionized water was used in the same operation as the product for cleaning and balancing to make the electrode signal response consistent with the initial test.

**1.5 Smart diagnosis of systematic CQAs covering biological, chemical, and physical CQAs for 30 batches of Xiao'er Xiaoji Zhike oral liquid in end-to-end real-world by information fusion**

The differences between inter-group samples and intra-group samples represented their discriminative power, and the ratio of inter-group variance and intra-group variance was used as a distinguishing index for different information methods.

$$SS_{inter}=\sum_{i=1}^{g} n_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})' \quad (5)$$

$$SS_{intra}=\sum_{i=1}^{g}\sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)' \quad (6)$$

$$\alpha=SS_{out}/SS_{in} \quad (7)$$

$$W_g=\alpha_g/\sum_{k=1}^{g}\alpha_g \quad (8)$$

$$UCL\left(T_A^2\right)_\alpha=\frac{(N-1)^2}{N}B_{(A/2),(N-A-1),\alpha} \quad (9)$$

$$\mathrm{UCL(SPE)}_\alpha=\theta_1\left[\frac{Z_\alpha\sqrt{2\theta_2 b_0^2}}{\theta_1}+1+\frac{\theta_2 b_0(b_0-1)}{\theta_1^2}\right]^{1/b_0} \quad (10)$$

Where $SS_{inter}$ is inter-group variance. $SS_{intra}$ is intra group variance. A is discriminative power. $W_g$ is weight. $UCL\left(T_A^2\right)_\alpha$ is D statistic Hotelling $T^2$. $UCL(SPE)_\alpha$ is the Square prediction error of Q statistics.

MSPC model Specific steps: the input was a matrix formed by the batch with the respective corresponding signal values, followed by normalization preprocessing, and leave-one-out cross-validation. The algorithm had a maximum number of iterations of 100 and singular value decomposition. Model outputs were Hotelling $T^2$ and F-Residuals. The resulting result, the D statistic (Hotelling $T^2$), represented how far the sample data were projected within the latent variable space from the origin of the latent variable space. The Q statistic squared prediction error (SPE) represented the orthogonal distance change of the sample data to the latent variable space of the principal component model.

**1.6 Mass transfer traceability of systematic CQAs for 30 batches of Xiao'er Xiaoji Zhike oral liquid in end-to-end real-world by multivariate process capability integrated with fuzzy mathematics**

**1.6.1 Fuzzy set theory is introduced to specify the specification of CQAs**

Firstly, the normal distribution of each quality attribute was tested. If the normal distribution is satisfied, the upper and lower limits of each quality attribute are introduced into the fuzzy set theory to make the upper and lower limits fuzzy by restricting the corresponding α-cut set, membership function, and indicator function.

Let R be the set of all real numbers and $F(R)=\{A|A:R\rightarrow[0,1]\ \text{A is a continuous function}\}$ be the set of all fuzzy sets on R.

$$\widetilde{USL}\ominus\widetilde{LSL}=\int_0^1 g(\alpha)(u_\alpha+l_\alpha)d\alpha \quad (11)$$

When the fuzzy specification is an indicator function $I_{\{x|x\geq LSL\}}$, $I_{\{x|x\leq USL\}}$; And $l_\alpha=LSL$,

$u_\alpha = USL$, for any $\alpha \in (0,1)$, there are $\widetilde{USL}_{\mathbb{L}} \ominus \widetilde{LSL}_{\mathbb{L}} = USL - LSL$, which stipulates that the membership functions of the upper fuzzy and the lower fuzzy are:

$$\widetilde{USL}_{\mathbb{L}}(x) = \begin{cases} 1, & x \le u_1 \\ (x-u_0)/(u_1-u_0), & u_1 < x < u_0 \\ 0, & u_0 \le x \end{cases} \tag{12}$$

$$\widetilde{LSL}_{\mathbb{L}}(x) = \begin{cases} 0, & x \le l_0 \\ (x-l_0)/(l_1-l_0), & l_0 < x < l_1 \\ 1, & l_1 \le x \end{cases} \tag{13}$$

$$\widetilde{USL}_L \ominus \widetilde{LSL}_L = \frac{1}{n+2}[(n+1)(u_1-l_1)+(u_0-l_0)] \tag{14}$$

Then the process capability index ($C_p$) can be generalized as a fuzzy $C_p$, as follows:

$$C_{\tilde{p}} = \frac{\widetilde{USL} \ominus \widetilde{LSL}}{6\sigma} = \frac{\frac{1}{n+2}[(n+1)(u_1-l_1)+(u_0-l_0)]}{6\sigma} \tag{15}$$

where, $\widetilde{USL} \in F(R)$, $\widetilde{USL}$ is the upper fuzzy specification limit and $F_U(R)$ represents the set of all upper fuzzy specifications, $\widetilde{LSL} \in F(R)$, $\widetilde{LSL}$ is the lower fuzzy specification limit and $F_L(R)$ represents the set of all lower fuzzy specifications. $u_0$, $l_0$ are any real numbers, $g \in [0,1]$, which is a non-increasing function, satisfying $g(0) = 0$, $\int_0^1 g(\alpha)d\alpha = 1$, For example, $g(\alpha) = (n+1)\alpha^n, n = 1,2,3,\ldots$ $\widetilde{USL} = (u_1, u_0)_{U\mathbb{L}}$ is the upper blur and $\widetilde{LSL} = (l_0, l_1)_{L\mathbb{L}}$ is the lower blur, when $g(\alpha) = (n+1)\alpha^n$, n=1,2,3…

**1.6.2 Mass transfer traceability of systematic CQAs for 30 batches of Xiao'er Xiaoji Zhike oral liquid in end-to-end real-world by multivariate process capability integrated with fuzzy mathematics**

The integrated weight assignment method, AHP-CRITIC, was implemented for further weight assignment of multivariate quality attributes of 90 batches of Xiaoer Xiaoji Zhike oral liquid from 3 manufacturing units. Firstly, an analytic hierarchy process model (AHP) was constructed according to the influencing factors and correlations contained in the three units of Xiaoer Xiaoji Zhike oral liquid. Next, a judgment matrix of the electronic tongue, electronic eye, pH, and chemical composition were divided into four categories by adopting the scale method of level 1 – 9 (**Table S3**).

Next, according to the formula (16-20), the weight vector and subjective assignment of the quality attributes of Xiaoer Xiaoji Zhike oral liquid, including the electronic tongue, electronic eye, pH, and chemical composition, were established by AHP.

$\Lambda_{\max}=\frac{1}{n}\sum_{i=1}^{n}\frac{(E\omega)_i}{\omega_i}$ (16)

$W=\sum_{j=1}^{n}\bar{a}_{ij}\ (i=1,2,\ldots,n)$ (17)

$W_i=W_i/\sum_{j=1}^{n}\bar{a}_{ij}\ (i=1,2,\ldots,n)$ (18)

CI =（$\lambda_{\max}$-$n$）/（$n$-1） (19)

CR =CI/RI (20)

where CI is the consistency index, λmax is the maximum characteristic root of the pairwise comparison matrix, and RI is the random consistency index, which is only related to the nth order of the matrix.

Based on this, the CRITIC method objectively assigns multiple evaluation indicators under the aforementioned four categories as the weight of the control layer. According to formula (21-24), the correlation coefficient matrix between various parameters was constructed by correlation analysis. Then the contrast strength and conflict were used to calculate the weight coefficient of the quality parameter weight coefficient, in which the contrast strength was displayed in the form of standard deviation.

$X_{ij}$'=$(x-x_{\text{mean}})/(x_{\max}-x_{\min})$ (21)

$C_j=\sigma_j\sum_{i=1}^{n}(1-r_{ij})$ ($j$=1，2，3…n) (22)

$\delta_j=\sum_{i=1}^{n}(1-r_{ij})$ (23)

$\omega_j=C_j/\sum_{j=1}^{n}C_j$ ($j$=1，2，3…n) (24)

where $x_{ij}$ is the value of the $j$th quality parameter of the $i$th sample, x is the measured value, and $x_{\text{mean}}$、$x_{\max}$、$x_{\min}$ are the mean, maximum, and minimum values of each quality parameter, respectively. $C_j$ is the influence degree of the $j$th quality parameter on the system, $\sigma_j$ is the standard deviation of the $j$th quality parameter, indicating the contrast strength $s_j$ of the quality parameter. $R_{ij}$ is the correlation coefficient between the $i$th and $j$th quality parameters. $\Delta_j$ represents the conflict between quality parameters. $\Omega_j$ is the objective weight of the $j$th quality parameter.

Moreover, the comprehensive weight of each quality attribute of the three production units of Xiaoer Xiaoji Zhike oral liquid is calculated according to formula (25), that is, the index weight obtained by the AHP method is multiplied by the index weight value obtained by the CRITIC method to calculate the comprehensive weight of each index.

$$w_c = \sqrt{w_{AHPij} w_{CRITICij}} \sum \sqrt{w_{AHPij} w_{CRITICij}} \quad (25)$$

Based on the results of fuzzy specification and comprehensive weight assignment, the fuzzy point estimation of the comprehensive quality digitization ability of intelligent manufacturing of three units in the production process of Xiaoer Xiaoji Zhike oral liquid was constructed. The fuzzy point estimation of the comprehensive quality digitization ability of intelligent manufacturing of three units was calculated by the sum of the products of the fuzzy process capability index of each quality attribute and its comprehensive weight. In the formula, n is taken as 1; $l_0 = \bar{x} - 2.58\sigma$, $l_1$ and $u_1$ are the minimum and maximum values of the data represented by each quality attribute, $l_0$~$l_1$ constitute the fuzzy lower limit interval and $u_1$~$u_0$ constitute the fuzzy upper limit interval. Finally, the fuzzy point estimation of the comprehensive quality digitization capability of intelligent manufacturing of three units is calculated by the sum of the product of the fuzzy process capability index of each quality attribute and its comprehensive weight.

**Table S3** AHP scale coefficients and their meanings for level 1 – 9

| Scale | Meaning |
|---|---|
| 1 | Indicate that the index *i* is as important as the index *j* |
| 3 | Indicate that index *i* is slightly more important than index *j* |
| 5 | Indicate that the index *i* is significantly more important than the index *j* |
| 7 | Indicate that index *i* is strongly more important than index *j* |
| 9 | Indicate that the index *i* is extremely important than the index *j* |
| 2，4，6，8 | The comparison of the importance of the two indicators is between the median of the above degrees |
| Reciprocal | The comparison of index j and index $A_{ji}$=1/$A_{ij}$ |

The fuzzy point estimation of the comprehensive quality digitization ability of intelligent manufacturing of three units was calculated by the sum of the products of the fuzzy $C_p$ of each quality attribute and its comprehensive weight. In the formula, n is taken as 1; $l_0 = \bar{x} - 2.58\sigma$, $l_1$, and $u_1$ are the minima and maximum values of the data represented by each quality attribute, $l_0$~$l_1$ constitute the fuzzy lower limit interval and $u_1$~$u_0$ constitute the fuzzy upper limit interval. Finally, the fuzzy point estimation of the comprehensive quality digitization capability of intelligent manufacturing of three units is calculated by the sum of the product of the fuzzy $C_p$ of each quality attribute and its comprehensive weight.

## 2. Results and discussion

### 2.1 Biological CQAs digitization and smart diagnosis for 30 batches of Xiao'er Xiaoji Zhike oral liquid in end-to-end real world by MIF-HEMT biosensor integrated UPLC-MS/MS

The lack of medication in children is a worldwide problem, and cough is a frequent and frequent disease in children. Authoritative data show that pediatric antitussive and expectorant drugs dominate the market of hospital proprietary Chinese medicine for children with a share of 41.71%. Xiao'er Xiaoji Zhike oral liquid, as a kind of Chinese patent medicine, fills the gap in the treatment of infantile food accumulation cough in China. According to the 2016 China Medical Statistics Annual Report of the Ministry of industry and information technology, Xiaoer Xiaoji Zhike oral liquid ranks first in the sales of Chinese patent medicine oral liquid preparations for relieving cough and resolving phlegm for children in China. Its total sales exceeded 970 million yuan for three consecutive years, covering 29 provinces and cities and nearly 1000 tertiary hospitals. The components of Xiaoer Xiaoji Zhike oral liquid are significant for screening high-quality candidates for productive cough due to indigestion. The components of Xiaoer Xiaoji Zhike oral liquid were identified by UPLC-MS/MS. 92 components were identified from Xiaoer Xiaoji Zhike oral liquid and the results were shown in **Table S4.**

**Table S4.** Ninety-two components identified from Xiaoer Xiaoji Zhike oral liquid by UPLC-MS/MS

| No. | Rt (min) | Molecular formula | Ion form | Theoretical molecular weight | Measured molecular weight | Δ (ppm) | MS2/MS3 | Identification conclusion |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.95 | $C_6H_{14}O_6$ | $[M-H]^-$ | 181.07066 | 181.07135 | 3.79 | 162.919(100),100.77(99.49),88.69(80.82),118.86(58.80),130.88(31.32) | L-Rhamnose monohydrate/ Mannitol |
| 2 | 1.07 | $C_{19}H_{34}O_{17}$ | $[M-H]^-$ | 533.17122 | 533.17065 | -1.08 | 190.89(100) | Glucopyranosyl fructofuranoside quinic acid |
| 3 | 1.07 | $C_7H_{13}NO_2$ | $[M+H]^+$ | 144.10190 | 144.10144 | -3.23 | 143.81(100),83.74(28.27),57.76(15.31),101.87(12.11) | Methyl piperidine-3-carboxylate3-piperidine formate |
| 4 | 1.15 | $C_{12}H_{22}O_{11}$ | $[M-H]^-$ | 341.10784 | 341.10739 | -1.31 | 178.98(100),16 | Sucrose |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | 0.88(21.46),112.84(21.00),142.94(19.66),118.89(19.05) | |
| 5 | 1.16 | $C_6H_6O_3$ | $[M+H]^+$ | 127.03897 | 127.03867 | -2.36 | 108.75(100),68.72(44.24),98.83(40.29),96.73(12.76) | 4-hydroxy-6-methyl-2-pyrone isomers |
| 6 | 1.24 | $C_4H_6O_5$ | $[M-H]^-$ | 133.01315 | 133.01421 | 7.97 | 114.77(100),132.88(3.40) | malic acid |
| 7 | 1.27 | $C_6H_{14}O_6$ | $[M-H]^-$ | 181.07066 | 181.07129 | 3.46 | | L-rhamnose monohydrate /D-Mannose |
| 8 | 1.34 | $C_5H_{11}NO_2$ | $[M+H]^+$ | 118.08625 | 118.08585 | -3.43 | 71.79(100) | Valine L-valine |
| 9 | 1.5 | $C_4H_5N_3O$ | $[M+H]^+$ | 112.05054 | 112.05029 | -2.22 | 111.83(100),69.74(79.12),83.89(11.86) | cytosine |
| 10 | 1.5 | $C_5H_5N_5$ | $[M+H]^+$ | 136.06177 | 136.06148 | -2.14 | | adenine |
| 11 | 2.03 | $C_5H_7NO_3$ | $[M-H]^-$ | 128.03422 | 128.03548 | 9.85 | 127.88(100),83.80(15.46) | Pyroglutamic acid |
| | | | $[M+H]^+$ | 130.04987 | 130.0495 | -2.84 | 83.70(100),101.83(2.34) | |
| 12 | 2.47 | $C_6H_6O_3$ | $[M-H]^-$ | 125.02332 | 125.02449 | 9.36 | 82.89(100),124.83(17.77),56.81(17.09) | 4-hydroxy-6-methyl-2-pyrone |
| | | | $[M+H]^+$ | 127.03897 | 127.0386 | -3.23 | 108.77(100),126.80(35.24),98.74(25.35),68.75(12.71) | |
| 13 | 2.54 | $C_6H_{13}NO_2$ | $[M+H]^+$ | 132.10190 | 132.10158 | -2.46 | 85.82(100) | leucine |
| 14 | 2.6 | $C_{14}H_{20}O_8$ | $[M-H]^-$ | 315.10744 | 315.10764 | 0.62 | | 4-[2-( β- D-glucopyranosyloxy) ethyl]– - 4-hydroxy-2,5-cyclohexadiene |
| 15 | 2.75 | $C_6H_{13}NO_2$ | $[M+H]^+$ | 132.10190 | 132.10149 | -3.14 | 85.85(100) | isoleucine |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 16 | 2.75 | $C_9H_{11}NO_3$ | $[M+H]^+$ | 182.08117 | 182.08066 | -2.80 | 164.90(100),135.87(21.71) | Tyrosine L-tyrosine |
| 17 | 2.98 | $C_{14}H_{20}O_8$ | $[M-H]^-$ | 315.10744 | 315.10764 | 0.62 | 178.92(100),118.81(70.43),142.98(64.95),112.84(59.37),152.88(47.00),161.01(40.99) | 4-[2-( β- D-glucopyranosyloxy) ethyl]– - 4-hydroxy-2,5-cyclohexadiene-1 |
| 18 | 3.93 | $C_9H_{13}NO$ | $[M+H]^+$ | 152.10699 | 152.10658 | -2.70 | 120.81(100) | N-methyltyrosyl |
| 19 | 4.45 | $C_{10}H_{13}N_5O_4$ | $[M+H]^+$ | 268.10403 | 268.10364 | -1.45 | 135.89(100) | Adenosine |
| 20 | 4.84 | $C_5H_4N_4O$ | $[M+H]^+$ | 137.04579 | 137.04564 | -1.07 | 108.85(100),80.71(67.21),118.80(12.03),136.90(10.19) | Hypoxanthine |
| 21 | 4.85 | $C_{14}H_{20}O_8$ | $[M-H]^-$ | 315.10744 | 315.10748 | 0.12 | 178.94(100),118.88(71.38),142.96(59.03),112.95(52.33),152.90(37.45),100.78(37.45),160.81(34.67) | Cornoside |
| 22 | 5.04 | $C_6H_9NOS$ | $[M+H]^+$ | 144.04776 | 144.04738 | -2.64 | 79.73(100),103.80(98.06),143.85(68.12), 52.73(34.79), 86.67(33.78), 125.77(19.49) –€) – 5 - (methylsulfonyl) pent-4-enenitrile5-methylsulfoxide pent-4-enenitrile | |
| 23 | 5.36 | $C_9H_{11}NO_2$ | $[M+H]^+$ | 166.08625 | 166.08604 | -1.29 | 119.88(100),148.80(3.96) | Phenylalanine |
| 24 | 6.2 | $C_{14}H_{20}O_8$ | $[M-H]^-$ | 315.10744 | 315.10773 | 0.91 | 178.82(100),118.72(66.49),112.89(64.49),14 | Kumamoside |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | 2.83(57.43),160.83(45.15),152.88(39.19),100.83(33.97) | |
| 25 | 6.5 | $C_9H_8O_3$ | $[M+H]^+$ | 165.05462 | 165.05414 | -2.91 | 120.71(100),136.92(11.49) | 4-Hydroxycinnamic acid |
| 26 | 7.14 | $C_{14}H_{20}O_8$ | $[M-H]^-$ | 315.10744 | 315.10751 | 0.21 | 134.81(100),152.89(16.80) | arbutin |
| 27 | 7.57 | $C_{16}H_{24}O_{10}$ | $[M-H]^-$ | 375.12857 | 375.12881 | 0.63 | 213.00(100),124.85(33.75),168.90(7.69),150.94(7.08) | Loganic acid |
| 28 | 7.82 | $C_{16}H_{18}O_9$ | $[M-H]^-$ | 353.08671 | 353.0867 | -0.02 | 190.99(100),178.87(44.24),134.87(9.01) | Neochlorogenic acid |
| 29 | 8.21 | $C_{19}H_{28}O_{12}$ | $[M-H]^-$ | 447.14970 | 447.14975 | 0.11 | 315.07(100),134.91(24.21),148.89(5.98) | Daphnetin B |
| 30 | 8.58 | $C_{20}H_{30}O_{12}$ | $[M-H]^-$ | 461.16535 | 461.16513 | -0.48 | 315(100),134.77(49.27),205.06(33.94),162.84(20.42)142.89(9.44) | Forsythoside E |
| 31 | 9.7 | $C_{16}H_{18}O_9$ | $[M-H]^-$ | 353.08671 | 353.08673 | 0.06 | 191.04(100),178.98(2.32) | Chlorogenic acid |
| | | | $[M+H]^+$ | 355.10236 | 355.1015 | -2.33 | 162.89(100),144.77(3.87) | |
| 32 | 9.89 | $C_{23}H_{32}O_{15}$ | $[M-H]^-$ | 547.16575 | 547.16547 | -0.50 | 205.08(100),222.97(83.84),367.19(45.60),189.83(30.81),341.23(15.27) | Z-sinapic acid Gentiobioside |
| 33 | 10.19 | $C_{21}H_{26}O_{13}$ | $[M+H]^+$ | 487.14462 | 487.14276 | -3.81 | 340.92(100),178.85(52.95) | 6-Methoxy-7-(6-O-β-D-xylopyranosyl- β-D-glucopyranosyloxy)-2H-1-benzopyran-2-one |
| 34 | 10.2 | $C_{16}H_{18}O_9$ | $[M-H]^-$ | 353.08671 | 353.08679 | 0.23 | 172.88(100),17 | Cryptochloro |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8.92(50.24),190.89(14.81),134.93(7.80) | genic acid |
| 35 | 10.68 | $C_{23}H_{32}O_{15}$ | $[M-H]^-$ | 547.16575 | 547.16547 | -0.50 | 204.87(100),222.99(48.14),367.11(44.27),189.90(23.51),341.11(9.63) | Isomers of z-sinapic acid Gentiobioside |
| 36 | 11.01 | $C_{21}H_{26}O_{13}$ | $[M-H]^-$ | 485.12897 | 485.12955 | 1.20 | 176.85(100),365.06(14.66),218.87(7.80) | 6-Methoxy-7-(6-O-β-D-xylopyranosyl- β-D-glucopyranosyloxy)-2H-1-benzopyran-2-one |
| | | | $[M+H]^+$ | 487.14462 | 487.1428 | -3.69 | 178.90(100),341.01(84.69),441.09(7.94) | |
| 37 | 11.23 | $C_{16}H_{24}NO_5^+$ | $[M+H]^+$ | 310.16490 | 310.16412 | -2.51 | 251.05(100) | Sinapine |
| 38 | 11.33 | $C_{27}H_{30}O_{16}$ | $[M+H]^+$ | 611.16066 | 611.15936 | -2.13 | | Kaempferol-3,7-diglucoside |
| 39 | 11.92 | $C_7H_{12}O_6$ | $[M-H]^-$ | 191.05501 | 191.05566 | 3.38 | 126.83(100),84.84(87.21),172.96(67.16),92.89(65.09),110.90(41.58),170.96(28.50) | Quinic acid |
| 40 | 12.49 | $C_{23}H_{32}O_{15}$ | $[M-H]^-$ | 547.16575 | 547.16565 | -0.17 | 204.93(100),367.13(69.59),222.99(65.62),189.84(28.21),341.07(9.49) | E-sinapic acid Gentiobioside |
| 41 | 12.6 | $C_{27}H_{30}O_{15}$ | $[M-H]^-$ | 593.15010 | 593.1499 | -0.33 | 371.19(100),532.99(4.46) | Naringin |
| | | | $[M+H]^+$ | 595.16575 | 595.1639 | -3.03 | 577.07(100),457.15(56.09),559.14(32.30),529.10(24.93),511.20(17.14),47 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 42 | 12.83 | $C_{23}H_{32}O_{15}$ | $[M-H]^-$ | 547.16575 | 547.16559 | -0.28 | 5.19(14.49),541.05(13.44) 223.01(100),205.01(66.35),264.99(49.67),367.10(39.80),295.09(38.84),325.05(31.05),189.86(17.43),384.99(13.61) | Isomers of e-sinapic acid Gentiobioside |
| 43 | 15.41 | $C_{28}H_{34}O_{15}$ | $[M-H]^-$ | 609.18140 | 609.1803 | -1.80 | 447.13(100),429.15(12.25),315.13(5.28) | Forsythoside J |
| 44 | 16.09 | $C_{15}H_{12}O_5$ | $[M+H]^+$ | 273.07575 | 273.07513 | -2.27 | 152.84(100),146.84(63.45),178.85(6.43) | Naringenin |
| 45 | 16.2 | $C_{29}H_{36}O_{15}$ | $[M-H]^-$ | 623.19705 | 623.19598 | -1.71 | 461.22(100),443.26(17.22),477.18(3.61) | Isomers of calycosin |
| 46 | 16.25 | $C_{27}H_{32}O_{15}$ | $[M-H]^-$ | 595.16575 | 595.16498 | -1.29 | 287.06(100),459.20(1.57) | Eriodictyosin |
| | | | $[M+H]^+$ | 597.18140 | 597.1796 | -2.96 | | |
| 47 | 16.47 | $C_{27}H_{30}O_{16}$ | $[M-H]^-$ | 609.14501 | 609.14478 | -0.38 | 300.94(100),299994(55.48),447.10(26.77),343.10(11.53),270.96(11.05) | Rutin |
| 48 | 16.7 | $C_{28}H_{34}O_{15}$ | $[M-H]^-$ | 609.18140 | 609.18079 | -0.99 | 447.08(100),429.20(7.62),315.21(4.81) | Isomers of Forsythoside J |
| 49 | 16.8 | $C_{15}H_{10}O_7$ | $[M+H]^+$ | 303.04993 | 303.04935 | -1.91 | 256.97(100),228.92(87.70),164.81(58.66),285.05(51.37),246.94(26.84),136.79(21.45) | Quercetin |
| 50 | 16.83 | $C_{27}H_{30}O_{16}$ | $[M-H]^-$ | 609.14501 | 609.14581 | | 300.97(100),300.20(50.61),270.94(8.44) | Kaempferol-3-o-sophoroside |
| | | | | 611.16066 | 611.159 | -2.72 | 303.01(100),465.06(24.65) | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 51 | 17.21 | $C_{15}H_{12}O_6$ | $[M+H]^+$ | 289.07066 | 289.06995 | -2.47 | 162.79(100),152.88(49.66),178.93(35.52),271.02(21.81) | Eriodictyol |
| 52 | 17.21 | $C_{21}H_{22}O_{10}$ | $[M+H]^+$ | 435.12857 | 435.12711 | -3.36 | 399.07(100),417.08(74.07),355.05(28.16),331.06(17.35),263.02(16.98) | Cherryoside |
| 53 | 17.21 | $C_{27}H_{32}O_{15}$ | $[M-H]^-$ | 595.16575 | 595.16498 | -1.29 | 459.10(100),234.92(11.31),287.03(7.30) | Neoeriocitrin |
| | | | $[M+H]^+$ | 597.18140 | 597.1793 | -3.48 | 451.05(100),432.99(79.65),435.14(57.62),331.03(38.46),579.08(34.87),561.04(31.04) | |
| 54 | 17.36 | $C_{29}H_{36}O_{15}$ | $[M-H]^-$ | 623.19705 | 623.19574 | -2.10 | 461.08(100),443.25(8.70) | Calycosin |
| 55 | 17.58 | $C_{29}H_{36}O_{15}$ | $[M-H]^-$ | 623.19705 | 623.19586 | -1.90 | 461.20(100),443.14(15.28),477.22(6.41),487.12(2.69) | Isofraxidin |
| 56 | 17.72 | $C_{16}H_{14}O_6$ | $[M+H]^+$ | 303.08631 | 303.08551 | -2.65 | 176.87(100),178.90(47.98),152.85(28.17),285.00(15.70) | Hesperetin |
| 57 | 18 | $C_{27}H_{30}O_{15}$ | $[M-H]^-$ | 593.15010 | 593.14984 | -0.43 | 285.02(100),447.12(45.41),284.00(18.95),327.00(9.22) | Kaempferol-3-o-rutinoside |
| | | | $[M+H]^+$ | 595.16575 | 595.1637 | -3.44 | 286.92(100),449.05(6.79) | |
| 58 | 18.18 | $C_{23}H_{26}O_{11}$ | $[M-H]^-$ | 477.13914 | 477.13943 | 0.61 | 160.86(100),315.16(20.60),281.09(2.65) | Xylopentose B |
| 59 | 18.78 | $C_{27}H_{32}O_{14}$ | $[M-H]^-$ | 579.17083 | 579.16956 | -2.19 | 270.95(100),176.94(1.57),459.07(1.26) | Narirutin |
| | | | $[M+H]^+$ | 581.18648 | 581.1845 | -3.49 | 419.05(100),417.14(69.03),43 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 60 | 19.3 | $C_{27}H_{30}O_{14}$ | $[M-H]^-$ | 577.15518 | 577.15527 | 0.15 | 4.95(65.02),545.07(44.14),273.07(33.47)401.09(31.17),383.16(29.03),527.13(26.44)<br>269.02(100) | Kaempferitrin |
| | | | $[M+H]^+$ | 579.17083 | 579.1694 | -2.42 | 433.04(100),270.94(24.54) | |
| 61 | 19.65 | $C_{27}H_{32}O_{14}$ | $[M-H]^-$ | 579.17083 | 579.16956 | -2.19 | 459.07(100),270.92(44.39),234.99(14.85),313.12(14.54),357.02(6.27) | Naringin |
| | | | $[M+H]^+$ | 581.18648 | 581.1852 | -2.24 | 417.05(100),435.04(94.65),419.04(71.46),545.19(41.04),273.03(34.76),315.07(29.73),562.94(27.10) | |
| 62 | 19.79 | $C_{34}H_{42}O_{19}$ | $[M-H]^-$ | 753.22365 | 753.22266 | -1.32 | 298.91(100),284.02(52.66),607.24(22.55) | Diguanylate Gentiobioside and its isomers |
| 63 | 19.79 | $C_{34}H_{44}O_{19}$ | $[M-H]^-$ | 755.23930 | 755.23834 | -1.28 | 301.05(100),489.09(14.62) | Forsythoside B |
| 64 | 20.01 | $C_{27}H_{30}O_{14}$ | $[M-H]^-$ | 577.15518 | 577.15521 | 0.05 | 268.96(100),413.08(4.59),431.07(2.62) | Vitexin rhamnoside |
| | | | $[M+H]^+$ | 579.17083 | 579.1688 | -3.58 | 270.99(100),433.06(8.97) | |
| 65 | 20.32 | $C_{28}H_{34}O_{15}$ | $[M-H]^-$ | 609.18140 | 609.18066 | -1.21 | 300.99(100),286.00(2.22),242.03(1.49) | Hesperidin |
| | | | $[M+H]^+$ | 611.19705 | 611.1949 | -3.53 | 465.07(100),447.08(50.10),449.07(48.78),574.99(33.95),593.20(23.60),303.06(22.78) | |
| 66 | 20.38 | $C_{28}H_{32}O_{15}$ | $[M-H]^-$ | 607.16575 | 607.16492 | -1.36 | 299.08(100),28 | Geranylgeran |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | 3.95(36.49) | yl isomers |
| | | | $[M+H]^+$ | 609.18140 | 609.1797 | -2.80 | 609.18140 | |
| 67 | 20.44 | $C_{34}H_{42}O_{19}$ | $[M-H]^-$ | 753.22365 | 753.22278 | -1.16 | 547.12(100),529.10(43.40),367.10(11.03) | Diguanylate Gentiobioside and its isomers |
| 68 | 20.68 | $C_{28}H_{32}O_{15}$ | $[M-H]^-$ | 607.16575 | 607.1651 | -1.06 | 299.02(100),283.96(38.42) | Geranylgeranyl isomers |
| | | | $[M+H]^+$ | 609.18140 | 609.1798 | -2.70 | 301.03(100),463.05(8.40),285.98(6.83) | |
| 69 | 21.03 | $C_{28}H_{32}O_{15}$ | $[M-H]^-$ | 607.16575 | 607.16693 | 1.95 | | Geranylgeranyl |
| | | | $[M+H]^+$ | 609.18140 | 609.1796 | -2.90 | 301.01(100),463.06(8.21),285.99(7.19) | |
| 70 | 21.09 | $C_{28}H_{34}O_{15}$ | $[M-H]^-$ | 609.18140 | 609.17981 | -2.60 | 301.02(100),343.15(15.70),489.12(13.43),325.11(10.58) | Neohesperidin |
| | | | $[M+H]^+$ | 611.19705 | 611.1951 | -3.13 | 302.98(100),301.06(81.61),464.95(77.05),302.06(68.50),447.16(64.42),449.05(37.82),575.13(35.11) | |
| 71 | 21.32 | $C_{48}H_{68}O_{28}$ | $[M-H]^-$ | 1091.38134 | 1091.37976 | -1.44 | 733.23(100),445.23(24.77),375.11(20.92),571.26(14.00) | Forsydoitriside A |
| 72 | 22.26 | $C_{15}H_{16}O_4$ | $[M+H]^+$ | 261.11214 | 261.11166 | -1.82 | 242.94(100),188.95(73.39),176.99(7.79) | Hespereolides |
| 73 | 22.75 | $C_{28}H_{36}O_{13}$ | $[M-H]^-$ | 579.20722 | 579.20709 | -0.22 | 371.08(100),459.28(19.15),533.03(11.52) | Acanthoside B |
| 74 | 22.8 | $C_{21}H_{22}O_5$ | $[M+H]^+$ | 355.15400 | 355.15326 | -2.08 | 285.10(100),305.17(51.32),150.89(24.28),136.92(18.78) | Imipramine |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 75 | 22.86 | $C_{34}H_{42}O_{19}$ | $[M-H]^-$ | 753.22365 | 753.22235 | -1.73 | 547.29(100),609.35(63.09),529(13(49.16),367.11(20.79),608.74(18.61),717.27(12.07) | Diguanylate Gentiobioside and its isomers |
| 76 | 23.33 | $C_{25}H_{30}O_{12}$ | $[M-H]^-$ | 521.16535 | 521.16559 | | 315.09(100),359.06(12.66),297.10(11.99),162.83(10.24),477.22(7.86) | Suspenoidside B |
| 77 | 23.72 | $C_{28}H_{34}O_{15}$ | $[M+H]^+$ | 611.19705 | 611.19623 | -1.33 | 302.96(100),465.12(49.21),593.00(19.34) | Isomers of hesperidin |
| 78 | 23.77 | $C_{28}H_{36}O_{13}$ | $[M-H]^-$ | 579.20722 | 579.20679 | -0.74 | 371.16(100),532.81(27.42),206.94(2.02) | Acanthoside B |
| 79 | 23.8 | $C_{21}H_{22}O_5$ | $[M+H]^+$ | 355.15400 | 355.15314 | -2.42 | 285.04(100),305.07(43.54),136.86(27.75)150.96(26.10),306.12(20.41),231.06(14.42) | Isomers of Imipramine |
| 80 | 23.91 | $C_{34}H_{42}O_{19}$ | $[M-H]^-$ | 753.22365 | 753.22296 | -0.92 | 609.17(100),651.22(32.24),691.32(9.44) | Gentiana diglycoside and its isomers |
| 81 | 24.24 | $C_{28}H_{32}O_{15}$ | $[M-H]^-$ | 607.16575 | 607.16486 | -1.46 | 301.01(100),463.22(12.74) | Isomers of geranitin |
| 82 | 25.05 | $C_{28}H_{34}O_{14}$ | $[M-H]^-$ | 593.18648 | 593.1861 | -0.64 | 285.04(100),308.98(4.87) | Isosakuranetin-7-rutinoside (didymin) |
| 83 | 25.77 | $C_{28}H_{34}O_{14}$ | $[M-H]^-$ | 593.18648 | 593.18622 | -0.44 | 285.04(100),327.11(19.23),473.13(16.14),309.12(5.41) | Poncirin |
| 84 | 26.05 | $C_{52}H_{82}O_{25}$ | $[M-H]^-$ | 1105.50614 | 1105.50378 | -2.14 | 695.40(100),519.38(9.27) | |
| 85 | 26.5 | $C_{52}H_{84}O_{24}$ | $[M-H]^-$ | 1091.52688 | 1091.52441 | -2.26 | 681.31(100),663.25(8.07) | Deapi-platycodin D |
| 86 | 27.11 | $C_{52}H_{82}O_{25}$ | $[M-H]^-$ | 1105.50614 | 1105.50415 | | 1075.34(100),895.42(87.76),4 | Unknown components 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | 85.30(26.47) | from Platycodon grandiflorus |
| 87 | 27.35 | $C_{54}H_{84}O_{26}$ | $[M-H]^-$ | 1147.51671 | 1147.51416 | -2.22 | 1117.57(100),937.47(71.58),485.39(22.68),895.41(13.20 | Unknown components 2 from Platycodon grandiflorus |
| 88 | 27.91 | $C_{54}H_{84}O_{26}$ | $[M-H]^-$ | 1147.51671 | 1147.5144 | -2.01 | 1117.43(100),937.32(73.43),485.32(25.28),895.26(12.59) | Unknown components 3 from Platycodon grandiflorus |
| 89 | 28.14 | $C_{54}H_{84}O_{26}$ | $[M-H]^-$ | 1147.51671 | 1147.51453 | -1.90 | 1117.61(100),937.54(53.75,485.37(27.33),895.40(13.59) | Unknown components 4 from Platycodon grandiflorus |
| 90 | 30.2 | $C_{26}H_{30}O_8$ | $[M+H]^+$ | 471.20134 | 471.19992 | -3.02 | 425.19(100),367.20(37.60),427.20(35.83),409.20(22.97),383.34(17.82) | Limonin |
| 91 | 30.7 | $C_{21}H_{22}O_8$ | $[M+H]^+$ | 403.13874 | 403.1377 | -2.59 | 388.13(100),373.01(59.03),342.11(9.05),355.08(5.34) | Nobiletin |
| 92 | 31.85 | $C_9H_{10}O_3$ | $[M+H]^+$ | 167.07027 | 167.06996 | -1.86 | 94.84(100),122.84(37.35),148.93(35.99) | Paeonol |

Furthermore, biological CQAs digitization of Xiaoer Xiaoji Zhike oral liquid was identified by MIF-HEMT biosensor integrated with UPLC-MS/MS. According to the high-resolution mass spectrometry data, in the negative ion mode, the excimer ion peak of compound M-5 was 595.16589 [M–H] - and the retention time was 16.52min. It was speculated that the molecular formula was $C_{27}H_{32}O_{15}$, which was the same as that of Neoeriocitrin, and the deviation from the theoretical molecular weight was 0.241ppm. The ionic fragment of the compound includes m/z 459.05 [M-H-$C_8H_8O_2$]- and m/z 287.07 [M-H-$C_{12}H_{20}O_9$]-. The fragment information was consistent with that reported in the literature, so it was speculated that compound M-5 was Neoeriocitrin. The specific

cracking principle of other biological CQAs was summarized in Supplementary Material.

According to the high-resolution mass spectrometry data, the excimer ion peak of compounds M-3 was 609.14563 $[M-H]^-$. It was speculated that the molecular formula of this compound was $C_{27}H_{30}O_{16}$, which was the same as that of Rutin, and the deviation from the theoretical value was 1.016 ppm. Besides, its ionic fragments include m/z 447 21 $[M-H-C_6H_{10}O_5]^-$, m/z 300. 96 $[M-H-C_{12}H_{20}O_9]^-$, m/z 270. 93 $[M-H-C_{12}H_{20}O_9-CH_2O]^-$, which was consistent with Rutin reported in the literature, so it was speculated that compounds A-5 are Rutin. In the negative ion mode, the excimer ion peak of compound M-1 was 461.16595 [M-H]-. It was speculated that the molecular formula was $C_{20}H_{30}O_{12}$, which was the same as that of Forsythin E, and the deviation from the theoretical molecular weight was 1.295 ppm. The ionic fragment of the compound included M / z315 10 $[M-H-C_6H_{10}O_4]^-$, m/z205. 01 $[M-H-C_8H_{10}O_3-C_4H_6O_3]^-$, m/z162. 79 $[M-H-C_6H_{10}O_4-C_8H_{10}O_3]^-$, m/z134. 95 $[M-H-C_6H_{10}O_4-C_9H_8O_4]^-$, which was consistent with forsythin e reported in the literature, so it was speculated that compound M-1 was Forsythin E.

In the negative ion mode, the excimer ion peak of compound M-1 was 461.16595 [M-H]-. It was speculated that the molecular formula was $C_{20}H_{30}O_{12}$, which was the same as that of Forsythin E, and the deviation from the theoretical molecular weight was 1.295 ppm. The ionic fragment of the compound includes M / z315 10 $[M-H-C_6H_{10}O_4]$-, m/z205. 01 $[M-H-C_8H_{10}O_3-C_4H_6O_3]$-, m/z162. 79 $[M-H-C_6H_{10}O_4-C_8H_{10}O_3]$-, and m/z134. 95 $[M-H-C_6H_{10}O_4-C_9H_8O_4]^-$, which was consistent with forsythin e reported in the literature, so it was speculated that compound M-1 was Forsythin E.

In the negative ion mode, the excimer ion peak of compound m-5 was 595.16589 [M–H] - and the retention time was 16.52 min. It was speculated that the molecular formula is $C_{27}H_{32}O_{15}$, which was the same as that of Neoeriocitrin, and the deviation from the theoretical molecular weight was 0.241 ppm. The ionic fragment of the compound includes M/z459 05 $[M-H-C_8H_8O_2]^-$ and m/z 287.07 $[M-H-C_{12}H_{20}O_9]^-$. The fragment information was consistent with that reported in the literature, so it was speculated that compound m-5 was Neoeriocitrin.

In the negative ion mode, the excimer ion peak of compound M-6 was 595.16620 [M–H] - and the retention time was 15.41 min. It was speculated that the molecular formula was $C_{27}H_{32}O_{15}$, which was the same as that of compound m-5, and the deviation between the measured molecular weight and

the theoretical molecular weight was 0.762 ppm. The ionic fragments of the compound include M/Z 287.06 [m-h-$C_{12}H_{20}O_9$] -. The fragment information was consistent with that reported in the literature. Therefore, it was speculated that compound M-6 was shengcaoside and was isomeric with compound m-5.

In the negative ion mode, the excimer ion peaks of compounds M-7 and M-8 were 609.18237 [M–H] - and 609.18195 [M-H]-, respectively, and the retention time was 21.14 min and 22.57 min. It was speculated that the molecular formula of both compounds was $C_{28}H_{34}O_{15}$, which belong to the isomer, and the deviation from the theoretical molecular weight was 1.598ppm and 0.908ppm respectively. The ion fragment of compound M-7 includes M / Z 300.99 [m-h-c12h20–9] - and the ion fragment of compound M-8 includes M / z489 22 [M-H-$C_7H_4O_2$]-, m/z325. 08 [M-H-$C_{16}H_{12}O_5$]-, and m/z301. 01[M-H-$C_{12}H_{20}O_9$]-. The fragment information was consistent with hesperidin and neohesperidin reported in the literature, so it was speculated that compound M-7 was hesperidin and compound M-8 was Neohesperidin .

In the negative ion mode, the excimer ion peak of compound M-9 was 593.18774 [M–H] - and the retention time was 28.33 min. It is speculated that the molecular formula is $C_{28}H_{34}O_{14}$, which was the same as that of citrinin, and the deviation from the theoretical molecular weight was 2.121ppm. The ionic fragment of the compound includes M/z473 13 [M-H-$C_8H_8O$]- and m/z285.04 [M-H-$C_{12}H_{20}O_9$]-. The fragment information was consistent with that reported in the literature, so it was speculated that compound M-9 was Lycium.

To sum up, ten biological CQAs were identified and adopted to the quality control.

## 2.2 Chemical CQAs digitization by UPLC and smart diagnosis for batch-to-batch quantitative chemical CQAs

**Table S5** UPLC-DAD method validation

| Samples | Project | Forsythiaside E | Neoeriocitrin | Hesperidin | Neohesperidin |
|---|---|---|---|---|---|
| Alcohol-precipitation intermediates | Linear equation | $y = 2\times10^6x$ - 28152 | $y = 7\times10^6x$ - 97300 | $y = 7\times10^6x$ - 163730 | $y = 8\times10^6x$ - 208502 |
| | Linear range ($\mu g\cdot mL^{-1}$) | 47.66-1525.00 | 23.52-752.50 | 77.81-2490.00 | 77.81-2490.00 |
| | $R^2$ | 0.9998 | 0.9991 | 0.9999 | 0.9998 |
| | LOD ($\mu g\cdot mL^{-1}$) | 0.019 | 0.101 | 0.208 | 0.282 |
| | LOQ ($\mu g\cdot mL^{-1}$) | 0.063 | 0.337 | 0.693 | 0.940 |

| | | | | | |
|---|---|---|---|---|---|
| | Precision | 1.20% | 1.67% | 0.95% | 1.34% |
| | Repeatability | 0.65% | 0.88% | 1.75% | 1.91% |
| | Stability | 0.37% | 0.38% | 0.37% | 0.11% |
| | recovery | 1.24% | 1.38% | 1.02% | 0.88% |
| Water-precipitation intermediates | Linear equation | $y = 1\times10^6x$ - 7086 | $y = 6\times10^6x$ - 50707 | $y = 6\times10^6x$ - 103226 | $y = 7\times10^6x$ - 219571 |
| | $R^2$ | 0.9995 | 0.9991 | 0.9993 | 0.9998 |
| | Linear range($\mu g\cdot mL^{-1}$) | 12.20-390.00 | 156.30-500.00 | 31.30-1000.00 | 78.10-1500.00 |
| | LOD($\mu g\cdot mL^{-1}$) | 0.042 | 0.187 | 0.343 | 0.421 |
| | LOQ($\mu g\cdot mL^{-1}$) | 0.140 | 0.623 | 1.143 | 1.403 |
| | Precision | 0.56% | 1.75% | 0.34% | 0.31% |
| | Repeatability | 0.98% | 0.68% | 0.75% | 0.82% |
| | Stability | 0.62% | 0.88% | 0.28% | 0.38% |
| | Recovery | 1.08% | 1.55% | 0.97% | 1.06% |
| Products | Linear equation | $y = 2\times10^6x$ - 28152 | $y = 7\times10^6x$ - 97300 | $y = 7\times10^6x$ - 163730 | $y = 8\times10^6x$ - 208502 |
| | $R^2$ | 0.9998 | 0.9991 | 0.9999 | 0.9998 |
| | Linear range($\mu g\cdot mL^{-1}$) | 47.70-1525.00 | 23.50-752.50 | 77.80-2.4900 | 77.80-2490.00 |
| | LOD($\mu g\cdot mL^{-1}$) | 0.021 | 0.106 | 0.202 | 0.263 |
| | LOQ($\mu g\cdot mL^{-1}$) | 0.069 | 0.354 | 0.674 | 0.875 |
| | Precision | 1.20% | 1.67% | 0.95% | 1.34% |
| | Repeatability | 0.65% | 0.88% | 1.75% | 1.91% |
| | Stability | 0.37% | 0.38% | 0.37% | 0.11% |
| | Recovery | 1.24% | 1.38% | 1.02% | 0.88% |

**Table S6** Analytical method validation results for the fingerprint analysis

| Samples | Precision | Repeatability | Stability |
|---|---|---|---|
| Alcohol-precipitated intermediates | 0.991–0.997 | 0.986–0.999 | 0.995–0.998 |
| Water-precipitated intermediates | 0.995–0.998 | 0.985–0.994 | 0.987–0.991 |
| Products | 0.997–0.999 | 0.992–0.998 | 0.994–0.997 |

**Table S7** Similarities of chromatograms of 30 batches of Xiaoer Xiaoji Zhike oral liquid

| Batches | Numbers | Similarity | Batches | Numbers | Similarity |
|---|---|---|---|---|---|
| 130121164 | S1 | 0.999 | 130121179 | S16 | 0.999 |
| 130121165 | S2 | 0.966 | 130121180 | S17 | 0.998 |
| 130121166 | S3 | 0.998 | 130121181 | S18 | 0.999 |
| 130121167 | S4 | 0.999 | 130121184 | S19 | 0.999 |
| 130121168 | S5 | 0.993 | 130121185 | S20 | 0.999 |
| 130121169 | S6 | 0.998 | 130121187 | S21 | 0.999 |
| 130121170 | S7 | 0.999 | 130121188 | S22 | 0.995 |

| | | | | | |
|---|---|---|---|---|---|
| 130121171 | S8 | 0.999 | 130121189 | S23 | 0.999 |
| 130121172 | S9 | 0.999 | 130121190 | S24 | 0.998 |
| 130121173 | S10 | 0.999 | 130121192 | S25 | 0.999 |
| 130121173 | S11 | 0.999 | 130121224 | S26 | 0.998 |
| 130121174 | S12 | 0.999 | 130121225 | S27 | 0.998 |
| 130121175 | S13 | 0.998 | 130121226 | S28 | 0.999 |
| 130121177 | S14 | 0.999 | 130121227 | S29 | 0.998 |
| 130121178 | S15 | 0.995 | 130121228 | S30 | 0.999 |

### 2.3 Smart diagnosis of systematic CQAs covering biological, chemical, and physical CQAs for 30 batches of end-to-end real-world Xiaoer Xiaoji Zhike oral liquid by information fusion

Firstly, a feature-level fusion strategy was implemented and obtained a 30 × 20 matrix via feature extraction of biosensors and other five sensor data. Next, a smart diagnosis by MSPC was performed. It was obvious that there were five abnormal batches in the finished product, and there were six abnormal batches in the alcohol-precipitation intermediate and four abnormal batches in the water-precipitation intermediate respectively (**Table S8**).

**Table S8.** Smart diagnosis for 30 batches of end-to-end real-world Xiaoer Xiaoji Zhike oral liquid

| Samples | Multi-sensors | Abnormal batches | Voting score |
|---|---|---|---|
| Products | Biosensor | 130121181, 130121182 | 16.71% |
| | UPLC | 130121165, 130121224 | 15.56% |
| | NIR | No | 16.83% |
| | Electronic eye | 130121165, 130121179, 130121185, 130121192 | 16.53% |
| | Electronic tongue | 130121174 | 16.70% |
| | Feature-level fusion | 130121166, 130121228, 130121225, 130121170, 130121184 | 16.72% |
| Alcohol-precipitation intermediate | UPLC | 130121165, 130121185, 130121227 | 15.30% |
| | NIR | 130121174, 130121188, 130121189, 130121227 | 25.29% |
| | Electronic eye | 130121180, 130121188, 130121189, 130121228 | 20.79% |
| | Electronic tongue | 130121178, 130121192 | 14.54% |
| | Feature-level fusion | 130121192, 130121228, 130121185, 130121188, 130121174, 130121189 | 24.08% |
| Water-precipitation intermediate | UPLC | 130121175 | 26.89% |
| | NIR | 130121179, 130121181, 130121187, 130121227 | 15.92% |
| | Electronic eye | 130121180, 130121224, 130121228 | 36% |
| | Electronic tongue | 130121174 | 3.25% |
| | Feature-level fusion | 130121171, 130121180, 130121187, 130121225 | 18.31% |

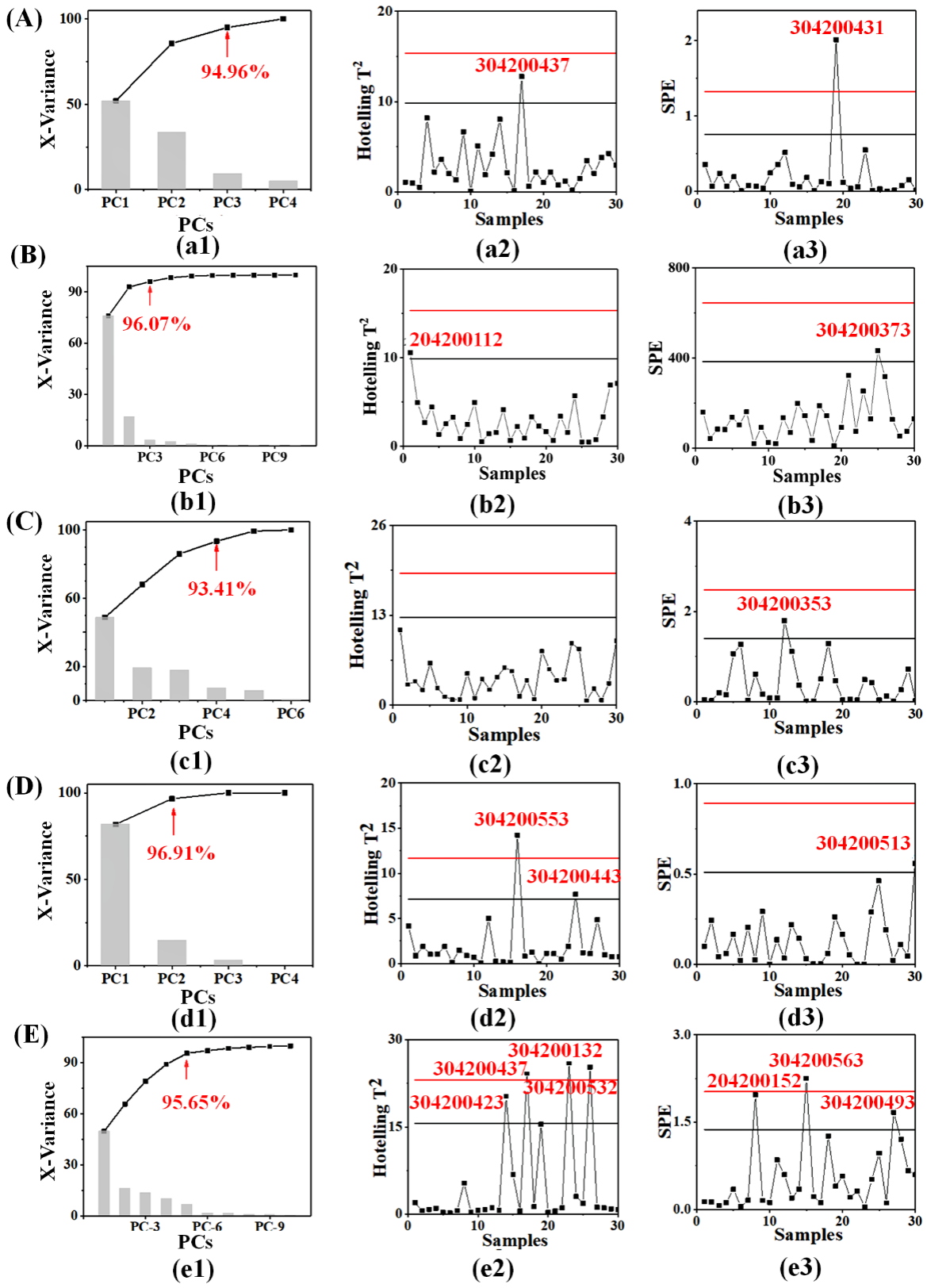


**Figure S1** Smart diagnosis for 30 verified batches of real-world Xiaoer Xiaoji Zhike oral liquid. (A) Smart diagnosis by UPLC. (a1) Cumulative principal component contribution diagram. (a2) Hotelling T2 control diagram.

(a3) SPE control diagram. (B) Smart diagnosis by near-infrared (NIR). (C) Smart diagnosis by the electronic tongue. (D) Smart diagnosis by electronic eye. (D) Smart diagnosis by feature-level information fusion.

**Table S9** Smart diagnosis for 30 verified batches of real-world Xiaoer Xiaoji Zhike oral liquid

| UPLC | NIR | E-eye | E-tongue | Feature-level fusion |
|---|---|---|---|---|
| 304200473<br>304200413 | 204200112<br>304200373 | 304200553<br>304200513 | 304200353 | 304200423，304200437，304200132，304200523，204200152，304200563，304200493 |

## 2.4 A novel end-to-end systematic CQAs traceability for 30 batches of end-to-end real-world Xiaoer Xiaoji Zhike oral liquid by multivariate process capability integrated with fuzzy mathematics

The problems were organized and hierarchical by analyzing the influencing factors and correlations contained in the three units of Xiaoer Xiaoji Zhike oral liquid. An analytic hierarchy process model was constructed, and the elements of each level were compared in pairs. The electronic tongue, electronic eye, pH, and chemical composition were divided into four categories by adopting the scale method of level 1 – 9, and the importance was compared with each other to assign score values to establish a judgment matrix (**Table S10**). Besides, the results of consistency test were λmax as 4.072, CI as 0.02401, and CR as 0.02668<0.10, indicating that random factors did not cause unreasonable weight vectors and meet the consistency requirements.

**Table S10.** The quality attribute judgment matrix of Xiaoer Xiaoji Zhike oral liquid

| | Electronic tongue | Electronic eye | pH | Composition | $W_i$ |
|---|---|---|---|---|---|
| Electronic tongue | 1 | 1/2 | 1/4 | 1/5 | 0.07689 |
| Electronic eye | 2 | 1 | 1/2 | 1/4 | 0.1332 |
| pH | 4 | 2 | 1 | 1/4 | 0.2322 |
| Composition | 5 | 3 | 4 | 1 | 0.5576 |

Furthermore, for the quality attributes of the alcohol precipitation unit, water precipitation unit and finished product unit in the production process of Xiaoer Xiaoji Zhike oral liquid, the objective assignment results of the CRITIC method are shown in the CRITIC in **Table S11**.

**Table S11** Comprehensive weight assignment of multiple quality attributes based on AHP-CRITIC method

| Weight | Indexs | Alcohol-precipitated intermediates | | Water-precipitated intermediates | | Finished products | |
|---|---|---|---|---|---|---|---|
| | | $W_{CRITIC}$ | $W_{Comprehensive}$ | $W_{CRITIC}$ | $W_{Comprehensive}$ | $W_{CRITIC}$ | $W_{Comprehensive}$ |
| Electronic tongue | P1 | 0.2071 | 0.0159 | 0.1338 | 0.0103 | 0.0990 | 0.0076 |
| | P2 | 0.2561 | 0.0197 | 0.1407 | 0.0108 | 0.1288 | 0.0099 |
| | P3 | 0.1159 | 0.0089 | 0.3180 | 0.0245 | 0.2086 | 0.0160 |
| | P4 | 0.1342 | 0.0103 | 0.1576 | 0.0121 | 0.1812 | 0.0139 |
| | P5 | 0.1410 | 0.0108 | 0.1207 | 0.0093 | 0.2347 | 0.0180 |
| | P6 | 0.1459 | 0.0112 | 0.1293 | 0.0099 | 0.1479 | 0.0114 |
| Electronic eye | L | 0.2457 | 0.0327 | 0.3933 | 0.0524 | 0.3141 | 0.0419 |
| | a | 0.5088 | 0.0678 | 0.2532 | 0.0337 | 0.3014 | 0.0402 |
| | b | 0.2455 | 0.0327 | 0.3535 | 0.0471 | 0.3845 | 0.0512 |
| pH | pH | 1.0000 | 0.2322 | 1.0000 | 0.2322 | 1.0000 | 0.2322 |
| Composition | Forsythoside E | 0.2026 | 0.1130 | 0.1799 | 0.1003 | 0.2769 | 0.1544 |
| | Neoeriocitrin | 0.1817 | 0.1013 | 0.2005 | 0.1118 | 0.1984 | 0.1107 |
| | Hesperidin | 0.4270 | 0.2381 | 0.2757 | 0.1537 | 0.3208 | 0.1789 |
| | Neohesperidin | 0.1887 | 0.1053 | 0.3440 | 0.1918 | 0.2039 | 0.1137 |