Special issue: External Quality Assessment in Laboratory Medicine

Original papers

Proficiency testing of maternal serum prenatal screening in second trimester in China, 2015

Xiaoyan Zhang, Wei Wang, Falin He, Kun Zhong, Shuai Yuan, Zhiguo Wang*

The first two authors in the by-line contributed equally to the work and thereby deserve joint first authorship.
National Center for Clinical Laboratories/Beijing Engineering Research Center of Laboratory Medicine, Beijing Hospital, Beijing, P.R. China

*Corresponding author: zgwang@nccl.org.cn

Abstract

Introduction: Prenatal screening and diagnosis is important for the detection of birth defects and genetic diseases. The nationwide proficiency testing (PT) of maternal serum prenatal screening in second trimester in China has been launched since 2003 and partly reflected the performance of screening laboratories. This study analysed the 2015 PT results to examine the performance of clinical laboratories and different platforms.

Materials and methods: Fifteen lyophilized samples with different concentrations divided into three panels, were distributed to 613 clinical laboratories in 2015. Acceptable performance was defined as scores more than 80% of acceptable responses with the evaluation criterion of $\pm$ 30%. The robust coefficient of variation (CV) was also analysed. Chi-square (χ^2) test was used to compare the acceptable performance while Kruskal-Wallis test and Mann-Whitney test were applied to compare the robust CV among analytes and mainstream platforms.

Results: Overall, 605, 61, 214, 416, 303 laboratories submitted effective results for alpha fetoprotein (AFP), total human chorionic gonadotropin (t-hCG), β-hCG, unconjugated estriol (uE3) and free β-hCG. The acceptable performances of AFP (μg/L), AFP (KIU/L), t-hCG, β-hCG, uE3, and free β-hCG were 98.45%, 99.24%, 95.58%, 98.72%, 94.50%, and 98.66%, respectively. The χ^2 test indicated significant differences existed in the acceptable performances among different analytes and platforms for uE3. Kruskal-Wallis test and Mann-Whitney test suggested the robust CV differed significantly in different analytes and platforms.

Conclusions: The majority of results were acceptable. However, further effort is needed to achieve the standardization and harmonization among analytes and various platforms, particularly for uE3.

Key words: maternal serum screening tests; proficiency testing; pregnancy trimester

Introduction

The high prevalence of birth defects and genetic diseases in China has seriously threatened the health condition of neonates and affected the quality of population (1). The objective of prenatal screening and diagnosis is to identify those women at the increased risk for an affected pregnancy and to maximize the options available to them (2). Maternal serum prenatal screening in second trimester is a screening test that collects the peripheral blood of pregnant women at 15-20 weeks (+ 6 days), combining the age-related risk of maternal for an affected pregnancy with the risks associated with the concentrations of biomarkers (3). The serum biomarkers involves alpha fetoprotein (AFP), total human chorionic gonadotropin (t-hCG), β-hCG, unconjugated estriol (uE3), free-β-hCG, and Inhibin-A (Inh-A), which have been used in combined forms to produce double (AFP and t-hCG/β-hCG/free β-hCG), triple (AFP, uE3, and t-hCG/β-hCG/free β-hCG) and quadruple (AFP, uE3, Inh-A, and t-hCG/β-hCG/free β-hCG) tests (4). Second trimester prenatal screening is economic, simple and non-invasive, and has been widely adopted since 1990s (5).

Prenatal screening has been started with imported software in China since 1990s (6). After two decades, most clinical laboratories and maternal and childcare service centres have provided prenatal screening services. To ensure the reliability of screening testing results in second trimester and assess the performance of laboratories simultaneously, the National Center for Clinical Laboratories (NCCL) in China has initiated the proficiency testing (PT) schemes for prenatal screening in second trimester nationwide since 2003, including AFP, hCG, β-hCG, free β-hCG and uE3. The frequency of PT evolves from once a year, twice a year to three times a year in 2015, shortening the monitoring period of institutions. The number of participants has increased from dozens to more than 600, covering 31 provinces nationwide (6). Information obtained from PT scheme can partly reflect the quality of screening laboratories by comparing its results with those of its peer group that have adopted the same platform (7). In addition, extensive results from the national PT may offer valuable information on the overall performance of prenatal screening laboratories within a country. This study presents the PT results of maternal serum prenatal screening in second trimester in 2015, in order to examine the performance of clinical laboratories and different platforms in China.

Materials and methods

Materials

The PT samples were commercial controls purchased from Baorong (Hangzhou, China) and prepared from human serum with additives of human or animal origin, chemicals, and stabilizers. All samples had been prepared, labelled and inspected to be non-reactive for the hepatitis B surface antigen (HBsAg), hepatitis C virus antibody (HCV) and human immunodeficiency virus antibody (HIV-1, HIV-2). All samples were provided in lyophilized form to increase the stability and would remain stable if stored integral at 2 to 8 °C until the expiration date. The homogeneity and stability of all samples were validated based on the China National Accreditation Service for Conformity Assess-

ment (CNAS) guidance CNAS-GL03 (8). In this survey, three PT test panels (20151, 20152, and 20153) consisting of fifteen samples were distributed to each participated laboratories in 2015, including low, normal, high and clinically important decision levels of analytes. Each sample of this PT scheme included five analytes: AFP, t-hCG, β-hCG, free β-hCG and uE3, respectively and was coded with six digits to facilitate analysis. The first four digits indicated the year, the fifth digit represented the lot of the panel, and the last digit stated the number of a sample in one panel.

PT program organization

In total, 613 laboratories in China were invited to participate in this survey for prenatal screening organized by NCCL in 2015. Fifteen control materials of three panels were assigned to participated laboratories in February 2015. Detailed instructions were provided to laboratories in hospital and maternal and child care service centres meanwhile, including details relating to the storage conditions, sample processing methods, and other procedures. Participants were required to handle the samples as guided and treat them equally as the patient specimens according to instructions. Participants were recommended to assay the first five samples (201511, 201512, 201513, 201514, 201515) in March, second five (201521, 201522, 201523, 201524, 201525) in July and the last five (201531, 201532, 201533, 201534, 201535) in October. Before testing, the lyophilized samples should be re-dissolved in 1mL of deionized or distilled water, and then placed upside-down with cover after standing at room temperature (18-25 °C) for 10 minutes. The results were submitted via the Clinet-EQA reporting system developed by NCCL (http://www.clinet.com.cn) before November 2015. Participants were expected to handle the samples using their routine methods to ensure that the results of this survey can reflect the actual ability of measuring.

Evaluation of the results

The participated laboratories were classified into several subgroups in terms of the platforms they adopted. For each analyte, we merely selected the

mainstream platforms with N ≥ 10 laboratories for preliminary data investigation. Overall, seven platforms were mentioned for all analytes in this study: Beckman (Brea, CA), Roche (Basel, Switzerland), Siemens DPC (München, Germany), PerkinElmer (Massachusetts, USA), Fenghua (Guangzhou, China), Darui (Guangzhou, China), Abbott (California, USA). The robust average of the results reported by all participants in a subgroup was considered as the assigned value, which was calculated using algorithm A introduced in ISO 13528 (9). For AFP, the result of each sample was considered acceptable if it fell in the range of ± 30% or 5 µg/L (whichever was larger) of the assigned value, for t-hCG, β-hCG, uE3, and free-β-hCG, the criterion was ± 30% established on the basis of the testing performance in China. As for other PT programmes, participants would obtain 20 points for an acceptable result. When 4 or 5 acceptable results for each panel (5 samples) were reported (80 or 100 points), the performance of this laboratory was determined to be satisfied. Unsatisfactory performance was attributed to scores below 80% for each analyte based on CLIA' 88 (10). The overall acceptable performance of each analyte was defined as (number of acceptable results) / (overall number of effective results). The acceptable performance of each panel was calculated as the ratio of the number of laboratories with satisfactory performance of this panel divided by the total number of laboratories of this panel. The acceptable performance of each platform was equivalent to the total number of laboratories with satisfactory performance of this platform divided by the total number of laboratories of this platform.

Statistical analysis

Data submitted by participants were calculated and statistic analysed *via* Microsoft Excel 2010 (Microsoft Inc., Redmond, Washington DC, USA), SPSS 19.0 (SPSS Inc., Chicago, IL, USA) and Clinet-EQA evaluation system designed by NCCL and developed by Clinet Information Technology (Beijing, China). For each sample, basic statistic parameters, such as the number of laboratories, arithmetic mean, standard deviation (SD), coefficient of variation (CV), robust average, robust standard deviation and robust CV were calculated and applied to assess the performance of screening laboratories. The parameters of each panel and platform were also analysed. To compare the acceptable performance among different analytes and various platforms, the chi-square (χ^2) test was used. The nonparametric Kruskal-Wallis (K-W) test and Mann-Whitney (M-W) test were also applied to identify significant differences of robust CV among various platforms and analytes. $P < 0.05$ was defined as the threshold of significance.

Results

In 2015, a total of 613 screening laboratories in hospitals and maternal and child health centres providing prenatal screening services were enrolled in this PT programme, in which 605 laboratories submitted effective results. Results of AFP were submitted by two different units, µg/L and KIU/L, respectively. Overall, 289, 316, 61, 214, 416, 303 laboratories submitted effective results for AFP (µg/L), AFP (KIU/L), t-hCG, β-hCG, uE3, and free-β-hCG. The overall acceptable performances of AFP (µg/L), AFP (KIU/L), t-hCG, β-hCG, uE3, and free β-hCG were 98.45%, 99.24%, 95.58%, 98.72%, 94.50%, and 98.66%, respectively. The results of each screening laboratories were scored and analysed in accordance with the criteria described above. Table 1 shows the acceptable performance of three panels for each analyte in 2015. For all analytes and panels, the proportion of laboratories with acceptable performance was above 90%, ranged from 92.8% (uE3, panel 20151) to 99.7% (AFP, KIU/L, panel 20153). The results of (χ^2) test suggested significant differences existed in the acceptable performance among different analytes ($P < 0.001$).

To further investigate the acceptable performance of different platforms, mainstream platforms with N ≥ 10 laboratories of each analyte were merely selected for data investigation (Table 2). There were three or two mainstream platforms for each of these biomarkers. The acceptable performance ranged from 93.0% (uE3, Beckman) to 100% (AFP: µg/L, DPC; t-hCG, Beckman; β-hCG, Abbott). The χ^2 test showed that the acceptable performance dif-

TABLE 1. Acceptable performance for three panels in 2015

Analyte (units)	Panel ID	Number of laboratories	Number of laboratories with acceptable performance	Acceptable performance (%)*
AFP (μg/L)	20151	289	282	97.58
	20152	272	269	98.90
	20153	276	273	98.91
AFP (KIU/L)	20151	316	312	98.73
	20152	310	308	99.35
	20153	300	299	99.67
t-hCG (IU/L)	20151	60	58	96.67
	20152	61	57	93.44
	20153	60	58	96.67
β-hCG (IU/L)	20151	214	210	98.13
	20152	205	203	99.02
	20153	208	206	99.03
uE3 (nmol/L)	20151	416	386	92.79
	20152	412	395	95.87
	20153	409	388	94.87
free-β-hCG (μg/L)	20151	303	298	98.35
	20152	300	297	99.00
	20153	292	288	98.63

AFP - alpha fetoprotein; t-hCG - total human chorionic gonadotropin; β-hCG - β-human chorionic gonadotropin; uE3 - unconjugated estriol; free-β-hCG - free β-human chorionic gonadotropin.
*The acceptable performance (%) of each panel was equivalent to the number of laboratories with acceptable performance of this panel divided by the total number of laboratories. Acceptable performance was attributed to scores ≥ 80. Acceptable result fell in the range of ± 30% or 5 μg/L (whichever was larger) of the assigned value, ± 30% for hCG, β-hCG, uE3, and free-β-hCG.

fered significantly among the mainstream platforms for uE3 ($P < 0.001$), but not for AFP (μg/L, $P = 1.000$), AFP (KIU/L, $P = 0.184$), t-hCG ($P = 1.000$), β-hCG ($P = 0.417$), and free β-hCG ($P = 0.183$).

The scatter diagram of robust CV of each sample for 6 analytes is shown in Figure 1. Each data point identified the robust CV of each sample (15 samples for each analyte). AFP and free-β-hCG showed better performance with robust CV below 10% while uE3 represented a poor performance with robust CV reached 30%. The results of Kruskal-Wallis test indicated statistical significant differences of robust CV from different analytes ($P < 0.001$).

To further evaluate the robust CV of different platforms in prenatal screening testing, Figure 2 shows the assigned value (robust average), robust SD and robust CV for each sample and each mainstream platform for AFP (μg/L), AFP (KIU/L), t-hCG, β-hCG, uE3, and free β-hCG. The samples in horizontal

axis were ordered by the increased concentration, and error bar represented the robust SD of each sample.
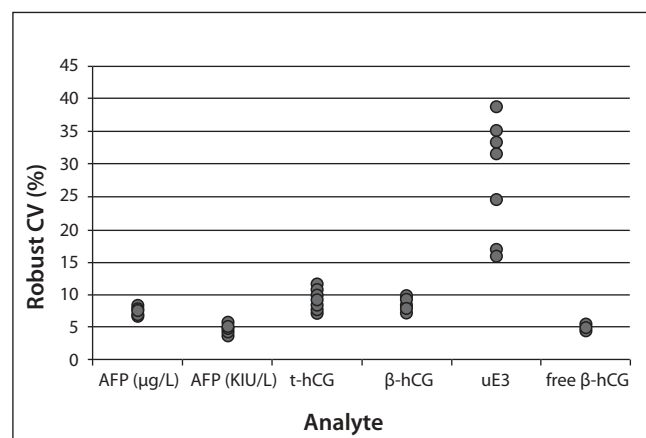

FIGURE 1. The robust CV of each PT sample for 6 analytes
AFP - alpha fetoprotein; t-hCG - total human chorionic gonadotropin; β-hCG - β-human chorionic gonadotropin; uE3 - unconjugated estriol; free-β-hCG - free β-human chorionic gonadotropin.

TABLE 2. Acceptable performance for each platform of three panels in 2015.

Analyte (units)	Platform	Number of panels reported effective results	Number of laboratories with acceptable performance	Acceptable performance (%)*
AFP (µg/L)	Beckman	491	488	99.39
	Roche	141	140	99.29
	DPC	66	66	100.00
AFP (KIU/L)	PerkinElmer	588	586	99.66
	Fenghua	178	177	99.44
	Darui	59	58	98.31
t-hCG (IU/L)	DPC	104	100	96.15
	Beckman	17	17	100.00
β-hCG (IU/L)	Beckman	466	463	99.36
	Roche	121	119	98.35
	Abbott	26	26	100.00
uE3 (nmol/L)	Beckman	497	462	92.96
	PerkinElmer	428	425	99.30
	DPC	97	95	97.94
free-β-hCG (µg/L)	PerkinElmer	589	587	99.66
	Fenghua	178	177	99.44
	Darui	59	58	98.31

AFP - alpha fetoprotein; t-hCG - total human chorionic gonadotropin; β-hCG - β-human chorionic gonadotropin; uE3 - unconjugated estriol; free-β-hCG - free β-human chorionic gonadotropin.
*The acceptable performance (%) of each platform was defined as the total number of laboratories with acceptable performance of this platform in three panels divided by the total number of laboratories. Acceptable performance was attributed to scores ≥ 80. Acceptable result fell in the range of ± 30% or 5 µg/L (whichever was larger) of the assigned value, ± 30% for hCG, β-hCG, uE3, and free-β-hCG.

For AFP (µg/L), the robust CV was higher in lower concentrations. For AFP (KIU/L), large fluctuations were seen in robust CV of Fenghua, while PerkinElmer had a preferable performance with robust CV below 4%. For t-hCG, the robust CV using Beckman was lower than that using DPC, except lots 201511, 201533, and 201534. The robust CV did not change drastically with the change of assigned value. For β-hCG, Abbott showed best among these three measurement systems. For uE3, the results indicated an observable decrease in robust CV along with the increased concentration. The robust CV was extremely large for uE3, however, the robust CV of PerkinElmer was relatively low with robust CV less than 10%. For free-β-hCG, the robust CV had larger degree of dispersion in lower concentrations among different platforms. The P values of Kruskal-Wallis test indicated the robust CV differed significantly among various platforms for AFP (µg/L, $P < 0.001$), AFP (KIU/L, $P < 0.001$), β-hCG ($P < 0.001$), uE3 ($P < 0.001$), and free-β-hCG ($P = 0.002$). Mann-Whitney test showed significant differences in robust CV between the two mainstream platforms for t-hCG ($P = 0.002$).

Discussion

Clinical laboratories desire to perform well and are required to participate PT schemes regularly by national standard and some local regulations in China. This report is an inaugural analysis of the national PT scheme for maternal serum prenatal screening in China. Information obtained from this PT programme might encourage participants to make effort to investigate the failures and improve the prenatal screening testing performance in Chi-
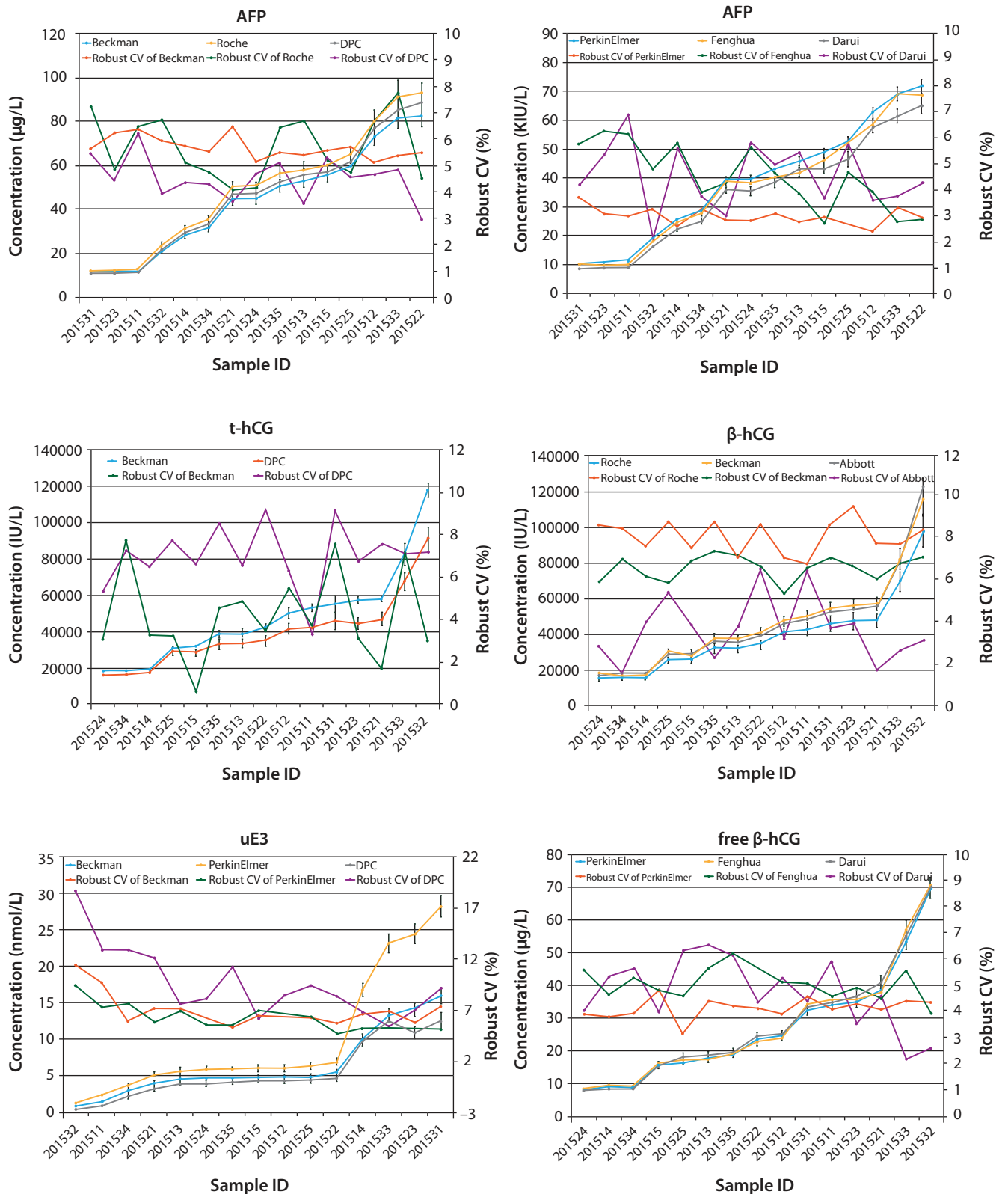
FIGURE 2. The robust average of concentration and robust CV of each PT sample of each platform
AFP - alpha fetoprotein; t-hCG - total human chorionic gonadotropin; β-hCG - β-human chorionic gonadotropin; uE3 - unconjugated estriol; free-β-hCG - free β-human chorionic gonadotropin.
The samples were ordered by the increased concentration, and error bar represented the robust SD of each sample.

na, which could help to the detection of birth defects and decrease the rates of birth defects ultimately.

A total of 605 laboratories in tertiary and secondary hospitals submitted effective results, covering mainstream platforms used nowadays. The numbers of laboratories participated in different measurements varied due to the disparity of selected screening protocols by laboratories (double, triple, or quadruple tests). Among them, AFP, β-hCG, free-β-hCG, and uE3 were customary chosen by laboratories while the number of laboratories using hCG was relative small (approximately selected by 10% laboratories). As the study for prenatal screening suggested that free-β-hCG was defined as an indicator with higher specificity in prenatal screening than hCG at 14 ~ 16 weeks during pregnancy (11).

College of American Pathologists (CAP) set the evaluation criterion as ± 3 standard deviations of the assigned value for AFP, t-hCG, uE3, and free-β-hCG. In our study, the acceptable criterion was defined as ± 30% or 5 μg/L (whichever was larger) of the assigned value for AFP, ± 30% for hCG, β-hCG, uE3, and free-β-hCG. The evaluation criterion was established based on the state-of-the-art performance in China, comprehensively considering the suggestions from extensive specialists of laboratory medicine and clinical medicine. Despite the criterion used in this study was different from the criteria in other PT schemes, it can certainly reveal the performance of laboratories in China.

The results of this study demonstrated that there was a significant difference in acceptable performance among different analytes, in which uE3 was comparatively lower. Study conducted by Lü discovered that the stability of uE3 was relatively worse than that of other measurements in prenatal screening (12). Likewise, our report showed the robust CV of uE3 was higher than that of other analytes, suggesting the results of samples for uE3 may be more possible to exceed the evaluation criterion, which contributed to the lower acceptable performance for uE3.

The acceptable performance of maternal serum prenatal screening in this study differed significant

among different platforms for uE3, but not for other analytes. The robust CV of uE3 using Beckman and PerkinElmer platforms was remarkably higher than that using DPC platform, suggesting that the dispersions of results using Beckman and PerkinElmer platforms were greater than using DPC platforms, thus corresponded to the lower acceptable performance. To further analyse the root causes, the different performance among various platforms for uE3 might be explained as the problems of methodology, instructions, practice, reporting, or even aware of quality control of laboratory staff. For AFP, t-hCG, β-hCG, and free β-hCG, although variations within various platforms generated discrepancies, and significant differences existed in the robust CV among different platforms, the acceptable performance had no significant statistical difference among those platforms.

A limitation of this study was the manufacture, transport, and storage technique of control materials, and simulated mature sera instead of samples of real pregnant women were used in the PT scheme, which may have caused the unavoidable matrix effect. There might be a significant difference among results of different platforms due to the matrix effect, so the assigned value (robust average) was also calculated by subgroups. In spite of this, the performance in the PT scheme could somewhat reflect the performance of daily practice in laboratories and platforms.

In conclusion, the results of this prenatal screening PT scheme indicated that the majority of results were acceptable in maternal serum prenatal screening in second trimester in China. However, significant difference existed in the acceptable performance among analytes and platforms for uE3. The PT scheme is vital, and further effort is needed to achieve the standardization and harmonization among various platforms, particularly for uE3.

Acknowledgements

the PT schemes. Clinet website (www.clinet.com.cn) is gratefully acknowledged for giving computer technology support to establish the network platform of data analysis.

Potential conflict of interest

None declared.

References

1. Lau TK, Fung HY, Rogers MS, Cheung KL. Racial variation in incidence of trisomy 21: survey of 57742 Chinese deliveries. Am J Med Genet 1998;75:386-8. https://doi.org/10.1002/(SICI)1096-8628(19980203)75:4<386::AID-AJMG7>3.0.CO;2-O.

2. Benn PA. Advances in prenatal screening for Down syndrome: I. General principles and second trimester testing. Clinica Chimica Acta 2002;323:1-16. https://doi.org/10.1016/S0009-8981(02)00186-9.

3. Xie Z, Lu S, Li H. Contingent triple-screening for Down's syndrome in the second trimester: a feasibility study in Mainland Chinese population. Prenat Diagn 2010; 30:74–6.

4. Wald NJ, Densem JW, George L, Muttukrishna S, Knight PG. Prenatal screening for Down's syndrome using inhibin-A as a serum marker. Prenat Diagn 1996;16:143-53. https://doi.org/10.1002/(SICI)1097-0223(199602)16:2<143::AID-PD825>3.0.CO;2-F.

5. Palomaki GE, Knight GJ, Lambert-Messerlian G, Canick JA, Haddow JE. Four years' experience with an interlaboratory comparison program involving first-trimester markers of Down syndrome. Arch Pathol Lab Med 2010;134:1685-91.

6. He F, Wang W, Zhong K, Wang Z. External quality assessment of maternal serum prenatal screening in second trimester during 2004-2013 in China. Chin J Lab Med 2014;37:270-5.

7. Xiao Y, Zhang C, Zhao H, Wang W, He F, Zhong K, et al. Status of External Quality Assessment on Tumor Markers in China. Clin Lab 2015;61:1383-90.

8. China National Accreditation Service for Conformity Assessment. Guidance on evaluating the homogeneity and stability of samples used for proficiency testing. CNAS-GL03. China: Beijing, 2006.

9. International Organization for Standardization. Statistical methods for use in proficiency testing by inter-laboratory comparisons. ISO 13528. Geneva: Switzerland, 2005.

10. Clinical Laboratory Improvement Amendments (CLIA). FDA, USA, 1988.

11. Shaw SW, Hsu JJ, Lee CN, Hsiao CH, Chen CP, Hsieh TT, et al. First- and second-trimester Down syndrome screening: current strategies and clinical guidelines. Taiwan J Obstet Gynecol 2008;47:157-62. https://doi.org/10.1016/S1028-4559(08)60073-X.

12. Lü S. Development of prenatal diagnosis technology and quality control in China. Chin J Lab Med 2010;33:101-5.