Clinical Biochemistry 63 (2019) 79–84

Contents lists available at ScienceDirect

Clinical Biochemistry

journal homepage: www.elsevier.com/locate/clinbiochem

Limitations of the Hoffmann method for establishing reference intervals using clinical laboratory data

Ying Zhang[a,b,1], Weibo Ma[c,1], Guocheng Wang[d], Yaqi Lv[a,b], Yaguang Peng[a,*], Xiaoxia Peng[a]

[a] Center for Clinical Epidemiology and Evidence-based Medicine, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China
[b] Department of Epidemiology and Biostatistics, School of Public Health, Capital Medical University, Beijing 100069, China
[c] Department of Obstetrics, Shunyi Maternal and Children's Hospital of Beijing Children's Hospital, Beijing 101300, China
[d] Department of Clinical Laboratory, Shunyi Maternal and Children's Hospital of Beijing Children's Hospital, Beijing 101300, China

ARTICLE INFO

Keywords:
Hoffmann method
Reference intervals
Thyroid hormones
Pregnancy
Acceptable deviation

ABSTRACT

Objective: To examine the effect of different limits of acceptability on the Hoffmann method for selecting "healthy" populations from laboratory test data.

Methods: Thyroid stimulating hormone (TSH) and free thyroxine (FT_4) were measured in Shunyi Maternal and Children's Hospital of Beijing Children's Hospital. The sample size of participants of TSH and FT_4 for reference intervals (RIs) establishment was 10,864 and 10,799, respectively. RIs were calculated by Hoffmann method with different acceptable deviations (α value). The validation data was collected prospectively and the out-of-range (OOR) values were calculated to examine the applicability of RIs with different acceptable deviations. The sample size for RIs validation was 880 and 867, respectively. The RIs were considered as valid when OOR was $< 10\%$.

Results: α value was set at different levels for establishing the RIs of TSH and FT_4. It was shown that the larger α value, the wider the RI. The RIs calculated by the Hoffmann method, under the default α value of 0.05, were much narrower than the previous findings. The OOR of both TSH and FT_4 were far more from 10% when the α value was 0.05. In this simulation, the OOR of TSH and FT_4 was not $< 10\%$ until the α value was set as 0.55 and 0.80, respectively. The established RIs were valid for both training dataset and validation dataset.

Conclusions: It can be somewhat subjective to define the acceptable deviation when selecting "normal" reference individuals. The default value of acceptable deviation may not be applicable in some cases. It is necessary to determine the acceptable deviation based on the certain condition, instead of using the default value directly.

1. Introduction

Reference intervals (RIs[2]) are regarded as the clinical decision-making tools and are widely used as relative standards to interpret laboratory test results [1,2]. RIs are generally defined as the interval from the lower reference limit to the upper reference limit, denoting the central 95% of values obtained from a healthy population [3]. The key step to establish RIs is to select a reference sample group, which is composed of "normal" or "healthy" reference individuals. However, it is difficult to define a reference individual as "normal" or "healthy", or to ensure the absence of subclinical issues. The methods of sampling for establishing RIs were based on two techniques: Direct and indirect sampling techniques [2]. The direct sampling method refers to establishing RIs using healthy individuals that are selected from reference populations with pre-specified inclusion and exclusion criteria. The result of direct sampling method may be precise and reliable; however, the recruitment of reference individuals is particularly challenging, costly and time consuming [4]. Therefore, the Clinical and Laboratory Standards Institute (CLSI) recommends the use of the indirect method when it proves difficult to collect samples from healthy subjects [2], for instance, pregnant women

Subsequently, researchers have paid more attention to the indirect

* Corresponding author at: Center for Clinical Epidemiology and Evidence-based Medicine, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, No.56 Nanlishi Road, Beijing 100045, China.
E-mail address: plwumi@hotmail.com (Y. Peng).

[1] Ying Zhang and Weibo Ma have equally contributions. They should be considered joint first author.

[2] Abbreviations: reference interval (RI), thyroid stimulating hormone (TSH), free thyroxine (FT4), out of range (OOR), standard deviation (SD), Clinical and Laboratory Standards Institute (CLSI), subclinical hypothyroidism (SCH).

https://doi.org/10.1016/j.clinbiochem.2018.11.005
Received 28 June 2018; Received in revised form 22 October 2018; Accepted 7 November 2018
Available online 09 November 2018

sampling technique, including Hoffmann method [5], Bhattacharya method [6], modified Bhattacharya method [7], and LIMIT method [8]. The Hoffmann method was first proposed in 1963 [5]; the obvious advantage to this method is that, compared with the direct method, hospital data is readily available without the recruitment of healthy individuals. Over the last few decades, the Hoffmann method has been widely used to establish RIs of different laboratory test indexes using sick/hospitalized population [9], including: erythrocyte count, stimulating thyroid hormone (TSH), mean corpuscular volume, hemoglobin, creatinine, calcium and ferritin [10–12]. Some RIs are proven to have a great agreement with those obtained on small cohorts of normal populations [9]. Therefore, the Hoffmann method is considered to be, to some extent, accurate and reproducible [12].

The Hoffmann method necessitates two assumptions: (1) hospital data for a particular index forms a Gaussian distribution; (2) the majority of measurements made in the hospital represent normal individuals [4]. The Hoffmann method is mainly based on the thought of Gaussian, plotting cumulative frequency graphs and regarding the linear portion of the cumulative graph as the "normal" population to calculate RIs [5]. Selecting "normal" reference individual is a key step to establish RIs. The "normal" population used to establish RI cannot be a population that was completely in the linear portion; otherwise it would lead to a narrow range. Therefore, to select a "normal" population, it is necessary to determine the range of data deviated from the linear portion, i.e. the acceptable deviation (α value). However, one of the biggest controversies of the Hoffmann method relates to its somewhat subjective selection of the α value when selecting "normal" reference individuals [10]. It is still universally undetermined the α value to obtain "normal" reference individuals. Therefore, further statistical simulations were conducted that aimed to examine the effect of different limits of acceptability in the Hoffman method on selecting the "healthy" population from laboratory test data. This would help us to reconsider the applicable conditions of Hoffmann method.

2. Materials and methods

2.1. Data collection and cleaning

The clinical laboratory data pertaining to thyroid stimulating hormone (TSH) and free thyroxine (FT_4) levels in the first trimester of pregnancy from July 1, 2013 to December 31, 2015 were collected from Shunyi Maternal and Children's Hospital of Beijing Children's Hospital. In addition, from January 1, 2018 to February 9, 2018, data of a completely new population were also collected from Shunyi Maternal and Children's Hospital of Beijing Children's Hospital for the purpose of validation. The exclusion criteria consisted of a history of thyroid disease or goiter, any other chronic diseases (liver, kidney, or other system diseases), nutrient deficiency, diabetes or hypertension in pregnancy, or twin pregnancies. TSH and FT_4 levels were measured using the Beckman UniCel DxI 800 Immunology Analyzer. This study was approved by the Ethics Committee of Beijing Children's Hospital, Beijing, China. We did not require formal consent for this type of study.

All further data cleaning processes for each index were conducted in training dataset and validation dataset. Firstly, missing values were checked and deleted. Missing values referred to missing data regarding id number, age or measurements. Secondly, any other record of the same patient was identified and deleted in order to avoid the presence of serial measurements. Finally, if the raw data showed Gaussian distribution, the outliers were excluded using the 3*SD* method [13,14] as follows: < >

$$y < \overline{x} \pm 3 \times SD \text{ or } y > \overline{x} \pm 3 \times SD$$

where y represents measurement, $\overline{x}$ is the mean of the test results, and SD represents the standard deviation.

Otherwise, Box-Cox transformation was applied to transform the data into normal distribution before outliers were deleted.

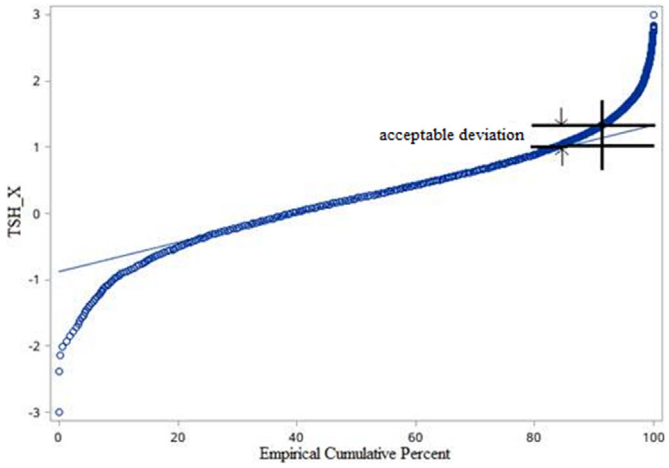

Fig. 1. A representative graph for TSH. Cumulative frequencies (circles) and regression line.

2.2. Selecting "normal" reference individuals using the Hoffmann method

Following data cleaning and Box-Cox transformation, the cumulative frequency for each test result was determined. The frequency of a test result (F_{Xi}) was calculated using the following formula [12]:

$$F_{X\,i} = (Count_{X\,i}/Count_{total}) \times 100\% \quad (1)$$

where $F_{X\,i}$ represents the frequency of a test result, $Count_{X\,i}$ represents the occurrence numbers of a test result in the data set, $Count_{total}$ represents the total number of results.

The cumulative frequency is $CF_{X\,i} = \sum_{m=2}^{i} F_{X\,m}$ ordered by X_i (2)

We plotted test value on the x-axis and cumulative frequency on the y-axis, which provided a visual analysis with a good approximation of the linear data so that the values from the linear portion of the cumulative frequency graph could be used. Subsequently, the "normal" population was determined using the acceptable deviation (α value), i.e. the range of data deviated from the linear portion. A representative graph with the acceptable deviation (α value) is shown in Fig. 1. However, the value of acceptable deviation was not fully determined ($\alpha = 0.05$ generally).

Therefore, we performed statistical simulations. The minimum deviation was set as 0.01, and the maximum deviation was determined when the percentage of test results within the RIs was greater than the 95%. Then, the portion of the data pool that was deemed to be "linear", i.e. the portion of the "normal" reference individual was selected according to different acceptable deviation.

2.3. Establishing and validating RIs

We calculated the 2.5th and the 97.5th percentiles of TSH and FT_4 levels with different alpha values using by bootstrapping at 1000 replications. If Box-Cox transformation was applied, RIs would be back-transformed to the original value. In order to compare the applicability of RIs calculated with different acceptable deviation e.g. alpha values, the OORs were calculated by "individual number of exceeding reference limits divided by total number of validation reference individuals" using completely new validation data. According to the CLSI document [2], the RIs may be considered valid if no more than 10% of the validation test results fall outside of the reference limits, which are presented by OORs in this study (the sum of the lower and upper OORs). In this study, data cleaning, selecting "normal" reference individuals and establishing RIs were performed using SAS version 9.4, while Box-Cox transformation was conducted on JMP version 13.0. The

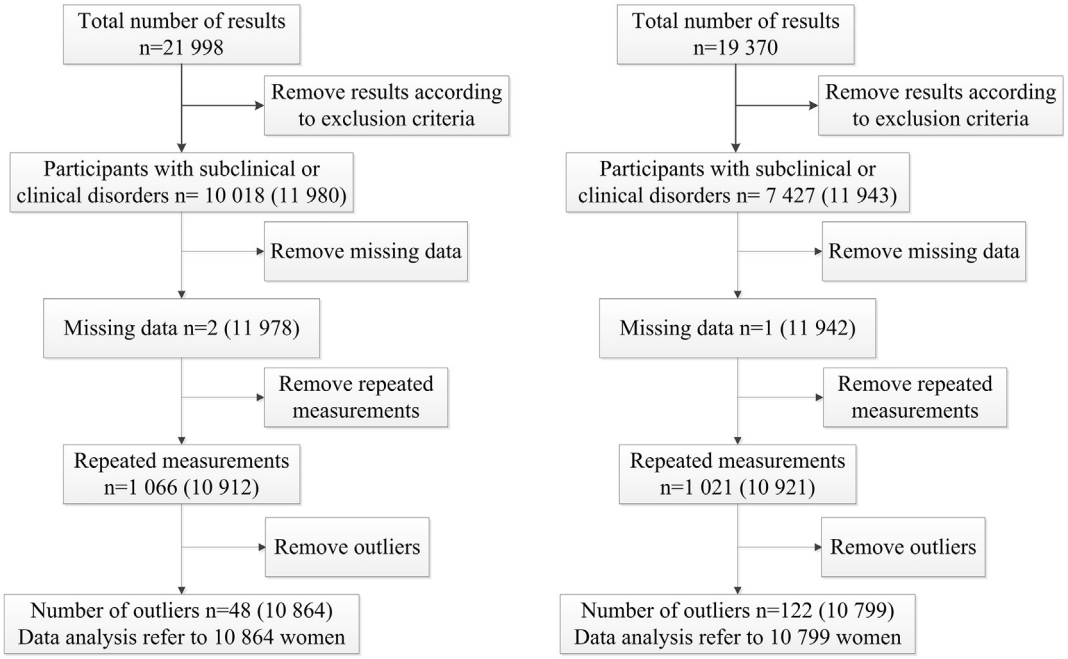

Fig. 2. Flow diagram of data cleaning.

statistical description process was implemented in the MedCalc Statistical version 15.10.0.

3. Results

3.1. Data cleaning and characteristics of data

10,018 and 7427 records of TSH and FT_4 were removed according to the exclusion criteria. All further data cleaning processes for each index is summarized in Fig. 2. After deleting missing data, repeated measurements and outliers, the sample size for TSH and FT_4 were 10,864 and 10,799, respectively. The sample size of validation data was 1038 and 1039 for TSH and FT_4, respectively. 97 and 112 records of TSH and FT_4 were removed according to the exclusion criteria. All further data cleaning processes of validation test results are displayed in Supplemental Fig. 1. Table 1 shows the characteristics of data after deleting outliers. Most participants were in the 26–30 year old group. The mean age of pregnant women was 27 years old. In addition, the characteristics of validation data are shown in Supplemental Table 1.

3.2. RIs at different α value

Neither the data of TSH nor FT_4 showed a Gaussian distribution. Therefore, Box-Cox transformation was applied (transformation parameter was 0.4 for TSH and − 0.6 for FT_4). We selected reference populations by the Hoffmann method under different α values. The RIs for TSH under different α values are given in Table 2. α values ranged from 0.01 to 0.65, while the corresponding percentage of normal participants ranged from 47.16% to 93.52%. The corresponding OOR values were calculated to compare the applicability of RIs calculated with different alpha values. OOR were no more than 10% until the α value equaled to 0.55. Similarly, the OOR of FT_4 was no more than 10% until the α value equaled to 0.80 (Table 3). α values ranged from 0.01 to 0.90 while the percentage of normal participants varied from 32.18% to 96.89%.

Fig. 3 depicts the portion of selected normal reference individuals at an α value of 0.05 (default value of acceptable deviation), and the portion at an α value of 0.55 and 0.80. The data indicated that too many reference individuals were excluded during selection of healthy reference individuals by the Hoffmann method when an α value equaled to 0.05, i.e. too many reference individuals were excluded when establishing RI of an index with a low disease prevalence.

Furthermore, the RIs calculated by Hoffmann method under an α value of 0.05 were much narrower than some published RIs for TSH and FT_4 in pregnancy (Supplemental Table 2).

Table 1
The characteristics of data after deleting outliers.

Age group	TSH[a] (mIU/L) (*n* = 10,864)				FT_4[b] (pmol/L) (n = 10,799)			
	n	$\bar{x}$	SD [c]	Median (Q_1,Q_3)	n	$\bar{x}$	SD	Median (Q_1,Q_3)
18– < 26 years	1835	1.43	0.97	1.27 (0.75, 1.90)	1829	10.87	1.98	10.66 (9.49, 11.96)
26– < 31 years	6058	1.42	1.01	1.23 (0.69, 1.92)	6025	10.95	1.94	10.79 (9.62–11.96)
31– < 36 years	2402	1.42	1.01	1.25 (0.69, 1.93)	2384	10.63	1.99	10.40 (9.36, 11.70)
36– < 41 years	512	1.52	1.03	1.32 (0.80, 2.01)	502	10.25	1.87	10.14 (8.97, 11.18)
41– < 46 years	54	1.67	1.10	1.48 (0.86, 2.10)	56	9.95	2.05	9.49 (8.58, 10.92)
46 years -	3	0.89	1.36	0.13 (0.09, 1.88)	3	10.49	2.43	9.49 (8.91, 12.32)

[a] TSH, thyroid stimulating hormone.
[b] FT_4, free thyroxine.
[c] SD: standard deviation.

Table 2
Reference intervals for TSH [a] and percentages of normal participants under different α values in validation data (n = 880).

α	Number of participants for establishing RI [b] (%)	RI (mIU/L)	Number of normal participants between RI (%)	OOR[c] lower	OOR upper
0.01	5950(54.77%)	0.74–2.07	415 (47.16%)	25.91%	26.93%
0.05[d]	6611(60.85%)	0.70–2.25	458 (52.05%)	24.32%	23.64%
0.10	7232(66.57%)	0.66–2.47	517 (58.75%)	22.73%	18.52%
0.15	7865(72.40%)	0.59–2.67	576 (65.45%)	18.86%	15.68%
0.20	8499(78.23%)	0.51–2.91	620 (70.45%)	16.93%	12.61%
0.25	8906(81.98%)	0.47–3.12	645 (73.30%)	16.02%	10.68%
0.30	9481(87.27%)	0.37–3.34	692 (78.64%)	12.73%	8.64%
0.35	9826(90.45%)	0.32–3.55	715 (81.25%)	11.14%	7.61%
0.40	10,021(92.24%)	0.28–3.67	729 (82.84%)	10.34%	6.82%
0.45	10,274(94.57%)	0.23–3.81	755 (85.80%)	8.52%	5.68%
0.50	10,428(95.99%)	0.18–3.88	775 (88.07%)	6.36%	5.57%
0.55[e]	**10,658(98.10%)**	**0.11–3.94**	**799 (90.80%)**	**4.20%**	**5.00%**
0.60	10,851(99.88%)	0.06–3.92	823 (93.52%)	1.36%	5.11%
0.65	10,864 (100.00%)	0.06–3.92	823 (93.52%)	1.36%	5.11%

The bold numbers represent the key point of OOR < 10%.
[a] TSH, thyroid stimulating hormone.
[b] RI, reference interval.
[c] OOR: proportions of out-of-range of RIs.
[d] α = 0.05, recommended by Hoffmann method.
[e] α = 0.55 so that OOR could be < 10%.

4. Discussion

In the present study, we studied the limitations of the Hoffmann method on selecting the "healthy" population from laboratory test data.

The Hoffmann method, in fact, is to some extent subjective because the acceptable deviation (α value) is determined by researchers. For example, α is equal to 0.05 in most studies [12]; however, too many reference individuals are excluded when the α value of 0.05 is used to select healthy reference individuals. The agreement between the "normal" reference individuals partitioned by the Hoffmann method, and the actual normal reference individuals, are impacted by the prevalence of diseases, which are diagnosed according to the index test results. The prevalence of thyroid dysfunction during pregnancy is relatively lower than other diseases. Shan ZY et al. reported that the prevalence of subclinical hypothyroidism (SCH) ranged from 4.53% to 6.15%, the prevalence of hypothyroxinemia ranged from 1.11% to 3.69%, and the prevalence of overt hypothyroidism ranged from 0.32% to 1.01% at 4 to 20 weeks of gestation [15]. The Chinese Society of Endocrinology claims that the prevalence of overt hypothyroidism is 0.3–0.5% in America and 1.0% in China [16]; Yang H et al. reported that the prevalence of SCH is 4.6% in the first trimester, the prevalence of overt hyperthyroidism is 0.2%, and the prevalence of subclinical hyperthyroidism is 0.6% [17]. In this case, a potential selective bias emerges from deleting more normal pregnancy data under an α value of 0.05. Furthermore, it indicates that in the first trimester, the RIs established by the Hoffmann method under the α value of 0.05 are narrower than some currently used RIs (Supplemental Table 2).

According to the CLSI, if no more than 10% of the test results fall

Table 3
Reference intervals for FT_4[a] and percentages of normal participants under different α values in validation data (n = 867).

α	Number of participants for establishing RI [b] (%)	RI (pmol/L)	Number of normal participants between RI (%)	OOR[c] lower	OOR upper
0.01	3669(33.98%)	9.36–10.79	279 (32.18%)	10.61%	57.21%
0.05[d]	7251(67.15%)	9.10–12.35	579 (66.78%)	7.50%	25.72%
0.10	7829(72.50%)	8.84–12.35	597 (68.86%)	5.42%	25.72%
0.15	8138(75.36%)	8.58–12.35	614 (70.82%)	3.46%	25.72%
0.20	9044(83.75%)	8.32–12.89	691 (79.70%)	2.19%	18.11%
0.25	9324(86.34%)	8.06–12.88	700 (80.74%)	1.15%	18.11%
0.30	9460(87.60%)	7.93–12.88	704 (81.20%)	0.69%	18.11%
0.35	9583(88.74%)	7.68–12.87	706 (81.43%)	0.46%	18.11%
0.40	9633(89.20%)	7.60–12.87	707 (81.55%)	0.35%	18.11%
0.45	9688(89.71%)	7.52–12.87	709 (81.78%)	0.12%	18.11%
0.50	9688(89.71%)	7.52–12.87	709 (81.78%)	0.12%	18.11%
0.55	9688(89.71%)	7.52–12.87	709 (81.78%)	0.12%	18.11%
0.60	9688(89.71%)	7.52–12.87	709 (81.78%)	0.12%	18.11%
0.65	9779(90.55%)	7.52–13.00	724 (83.51%)	0.12%	16.38%
0.70	9779(90.55%)	7.52–13.00	724 (83.51%)	0.12%	16.38%
0.75	9779(90.55%)	7.52–13.00	724 (83.51%)	0.12%	16.38%
0.80[e]	**10,419(96.48%)**	**7.54–13.98**	**793 (91.46%)**	**0.12%**	**8.42%**
0.85	10,688(98.97%)	7.54–14.77	832 (95.96%)	0.12%	3.92%
0.90	10,799(100.00%)	7.55–15.35	840 (96.89%)	0.35%	2.77%

The bold numbers represent the key point of OOR < 10%.
[a] FT_4, free thyroxine.
[b] RI, reference interval.
[c] OOR: proportions of out-of-range of RIs.
[d] α = 0.05, recommended by Hoffmann method.
[e] α = 0.80 so that OOR could be < 10%.

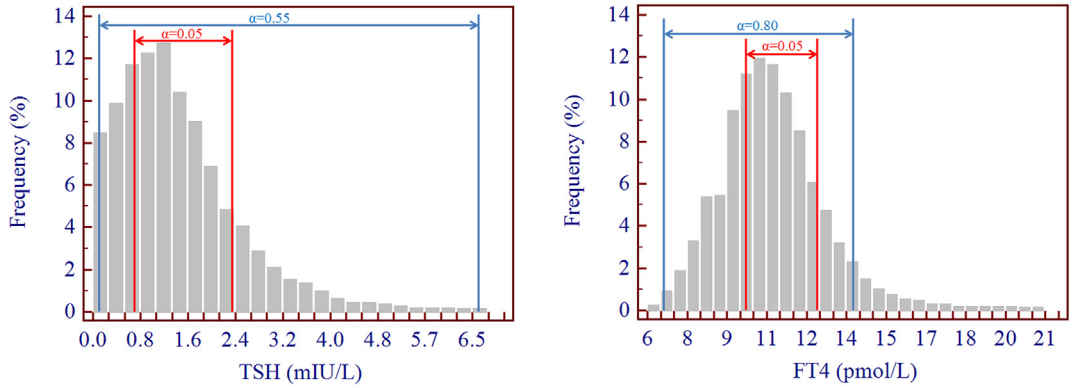

Fig. 3. The portion of reference individuals for establishing RIs of TSH and FT_4 at different α values. TSH, thyroid stimulating hormone. FT_4, free thyroxine.

outside the RIs, the RIs may be considered valid [2]. However, in this study, when the α value equaled to 0.05, the OOR of TSH (47.96%) and FT_4 (33.22%) was far greater than the prevalence of thyroid diseases during pregnancy, showing that the Hoffmann method maybe not applicable for some indexes with lower prevalence. We established RIs under different α values and we found that as α values increased, the RIs became wider and the percentages of normal participants were greater. Although the α value of 0.05 is a default deviation level used to establish RIs by the Hoffmann method, such α value lead to a very large OOR value in this study. As far as TSH goes, the OOR was > 10% until the α value was enlarged to 0.55. In terms of FT_4, the α value was equal to 0.80 when the OOR was < 10%. However, this may not ensure that the selected reference individuals from the clinical laboratory were all "normal" if the α value was too large.

Therefore, the default value of 0.05 may not be applicable in some indexes with low prevalence. It is necessary to determine the acceptable deviation based on the specific condition, for example, by verifying the RIs by conducting a pilot study rather than directly using the default value. In addition, there are also some other factors that influence the applicability of the Hoffmann method, for example, the data sources and the characteristics of study subjects [26]. The Hoffmann method requires that more than 50% of the reference individuals should be healthy [4].

There are some limitations in our study. Firstly, there were indeed likely differences in the expected values of TSH and FT_4 throughout pregnancy. However, the present study focused to examine the applicable conditions of Hoffman method rather than to establish the RIs of TSH and FT4 for pregnancy. Therefore, only the data of TSH and FT_4 of the first trimester was collected. In the future, more attention will be given to the calculation of RIs by trimester. Secondly, we did not assess the urinary iodine levels; however, iodine intake gradually increased among many inland areas of China over the past several decades [27] so that the iodine intake could be considered sufficient. Thirdly, although fasting is not mandatory for thyroid function testing, fasting related factors may influence the results of the study [12].

5. Conclusions

Although the Hoffmann method saves time and money, there are some limitations to establish reliable RIs using clinical laboratory data. It is somewhat subjective in defining the acceptable deviation when selecting "normal" reference individuals. The default value of acceptable deviation may not be applicable in some cases. Therefore, it is necessary to determine the acceptable deviation based on the specific conditions, for example, by verifying the RIs by conducting a pilot study, instead of using the default value directly.

Funding

This work was supported by National Health and Family Planning Commission of China [grant number WJW1601]; Based-clinical research collaboration Foundation of Capital Medical University, Beijing [grant number 16JL01]; and Beijing Municipal Administration of Hospitals' Youth Programme [grant number QML20171203].

Declarations of interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgements

We acknowledge all participants for their cooperation and sample contributions. We thank Mr. Ali Abbas for checking the English language and grammar of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinbiochem.2018.11.005.

References

[1] k. Adeli, V. Higgins, D. Seccombe, C.P. Balion, G. Cembrowski, A.A. Venner, J. Shaw, CSCC Reference Interval Harmonization (hRI) Working Group, National Survey of Adult and Pediatric Reference Intervals in Clinical Laboratories across Canada: a Report of the CSCC Working Group on Reference Interval Harmonization, Clin.Biochem. 50 (2017) 925–935, https://doi.org/10.1016/j.clinbiochem.2017.06.006.

[2] CLSI, Defining, Establishing and Verifying Reference Intervals in the Clinical Laboratory: Approved Guideline—Third Edition. CLSI Document C28-A3, Clinical and Laboratory Standards Institute, Wayne PA, 2010.

[3] D.A. Colantonio, L. Kyriakopoulou, M.K. Chan, C.H. Daly, D. Brinc, A.A. Venner, M.D. Pasic, D. Armbruster, K. Adeli, Closing the gaps in pediatric laboratory RIs: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children, Clin. Chem. 58 (2012) 854–868, https://doi.org/10.1373/clinchem.2011.177741.

[4] J.L. Shaw, A. Cohen, D. Konforte, T. Binesh-Marvasti, D.A. Colantonio, K. Adeli, Validity of establishing pediatric reference intervals based on hospital patient data: a comparison of the modified Hoffmann approach to CALIPER reference intervals obtained in healthy children, Clin. Biochem. 47 (2014) 166–172, https://doi.org/10.1016/j.clinbiochem.2013.11.008.

[5] R.G. Hoffmann, Statistics in the practice of medicine, JAMA 185 (1963) 864–873.

[6] C.G. Bhattacharya, A simple method of resolution of a distribution into Gaussian components, Biometrics 23 (1967) 115–135.

[7] H. Baadenhuijsen, J.C. Smith, Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya procedure, J. Clin. Chem. Clin. Biochem. 23 (1985) 829–839.

[8] S. Poole, L.F. Schroeder, N. Shah, An unsupervised learning method to identify reference intervals from a clinical database, J. Biomed. Inform. 59 (2016) 276–284, https://doi.org/10.1016/j.jbi.2015.12.010.

[9] O.P. Soldin, J.R. Dahlin, E.G. Gresham, J. King, S.J. Soldin, IMMULITE® 2000 age and sex-specific reference intervals for alpha fetoprotein, homocysteine, insulin,

insulin-like growth factor-1, insulin-like growth factor binding protein-3, C-peptide, immunoglobulin E and intact parathyroid hormone, Clin. Biochem. 41 (2008) 937–942, https://doi.org/10.1016/j.clinbiochem.2008.04.025.
[10] G.S. Feng, W.Q. Song, Y.J. Lü, X.L. Nie, S.Y. Cai, T.Y. Liu, X.X. Peng, The application of Hoffmann method in the establishment of children's reference intervals of erythrocyte count, Zhonghua Yi Xue Zazhi 96 (2016) 1742–1745, https://doi.org/10.3760/cma.j.issn.0376-2491.2016.22.006 (Article in Chinese).
[11] A. Katayev, J.K. Fleming, D. Luo, A.H. Fisher, T.M. Sharp, Reference intervals data mining no longer a probability paper method, Am. J. Clin. Pathol. 143 (2015) 134–142, https://doi.org/10.1309/ajcpqprnib54wfkj.
[12] A. Katayev, C. Balciza, D.W. Seccombe, Establishing reference intervals for clinical laboratory test results: is there a better way? Am. J. Clin. Pathol. 133 (2010) 180–186, https://doi.org/10.1309/ajcpn5bmtsf1cdyp.
[13] Q. Zhou, J. Xu, W. Xie, S. Li, X. Li, Use of robust ZB and ZW to evaluate proficiency testing data, Clin. Chim. Acta 412 (2011) 936–939, https://doi.org/10.1016/j.cca.2011.01.027.
[14] Q. Zhou, Z. Shen, S. Li, X. Li, W. Wang, Z. Wang, Robust and traditional statistical methods in the establishment of immunoglobulin E target values in external quality assessment program, Clin. Chim. Acta 387 (2008) 66–70.
[15] Z.Y. Shan, Y.Y. Chen, W.P. Teng, X.H. Yu, C.Y. Li, W.W. Zhou, B. Gao, J.R. Zhou, B. Ding, Y. Ma, Y. Wu, Q. Liu, H. Xu, W. Liu, J. Li, W.W. Wang, Y.B. Li, C.L. Fan, H. Wang, R. Guo, H.M. Zhang, A study for maternal thyroid hormone deficiency during the first half of pregnancy in China, Eur. J. Clin. Investig. 39 (2009) 37–42 (doi: 10.1111/j.1365-2362.2008.02055.x).
[16] Chinese Society of Endocrinology, Guidelines for the diagnosis and treatment of thyroid diseases in pregnancy and postpartum, Clin. J. Endocrinol. Metab. 28 (2012) 354–371 (Article in Chinese).
[17] H. Yang, M. Shao, L. Chen, Q. Chen, L. Yu, L. Cai, Z. Lin, C. Zhang, X. Lu, Screening strategies for thyroid disorders in the first and second trimester of pregnancy in China, PLoS ONE 9 (2014) e99611, , https://doi.org/10.1371/journal.pone.0099611.
[26] J.L. Shaw, T. Binesh Marvasti, D. Colantonio, K. Adeli, Pediatric reference intervals: challenges and recent initiatives, Crit. Rev. Clin. Lab. Sci. 50 (2013) 37–50, https://doi.org/10.3109/10408363.2013.786673.
[27] W. Zhao, C. Han, X. Shi, C. Xiong, J. Sun, Z. Shan, W. Teng, Prevalence of goiter and thyroid nodules before and after implementation of the universal salt iodization program in mainland China from 1985 to 2014: a systematic review and meta-analysis, PLoS ONE 9 (2014) e109549, , https://doi.org/10.1371/journal.pone.0109549.