Journal of Tianjin Medical University

Vol. 26, No. 4 Jul. 2020

文章编号 1006-8147(2020)04-0378-05

论著

利用 McNemar 方案及 ROC 曲线评价进口品牌甲功检测结果临床诊断价值

袁雪1,2,张柏畅2,李会强1

(1.天津医科大学医学检验学院,天津300070;2.天津医科大学第二医院输血科,天津300211)

摘要 目的:评估不同进口品牌罗氏 Cobas E602、西门子 Centaur XP 及贝克曼 DXI800 间血清甲状腺功能(甲功)检测项目临床诊断价值及诊断效能是否一致。方法:选择初诊甲状腺功能亢进症(甲亢)患者 63 例、初诊甲状腺功能减退症(甲减)患者 66 例及健康体检者 70 名作为对照组。采集 3 组未经治疗期血清标本,检测甲功 5 项(TSH、FT₃、T₄、T₃、T₄、T₃),采用多因素联合诊断受试者工作特征曲线(ROC 曲线)及 WS/T505-2017《定性测定性能评价指南》提供的 McNemar 方案评价诊断价值及效能。结果:多因素联合诊断 ROC 曲线分析,3 个品牌 5 个项目联合检测,甲亢患者的曲线下面积(AUC)在 0.98~0.99 之间,甲减在 0.81~0.84 之间,3 个品牌比较差异无统计学意义(均 P>0.05);根据 McNemar 方案,以各品牌既定临界点(参考范围上下限)作为判断点,甲亢患者 3 个品牌诊断效能基本一致,仅 TSH 项目西门子具备更好临床敏感性的同时降低了临床特异性,甲减患者 3 个品牌互有优势,但 FT₃及 FT₄ 无论是甲亢或甲减,3 个品牌临床敏感性及临床特异性均较低(13.70%~50.00%)。结论:Cobas E602、Centaur XP 及 DXI800 检测血清甲功项目时,对于甲亢的诊断价值均较高并基本一致,建议使用该 3 个品牌的检验机构可以开展甲亢检测结果的互信;对于甲减的诊断价值虽然中等并一致,但 FT₃及 FT₄ 3 个品牌临床敏感性均较低。

关键词 甲状腺功能检测;McNemar 方案; ROC 曲线;临床诊断价值

中图分类号 R446.1

文献标志码 A

Evaluation of the clinical diagnostic value of imported brand in the results of the thyroid function test by McNemar scheme and ROC curve

YUAN Xue^{1,2}, ZHANG Bai-chang², LI Hui-qiang¹

(1. School of Medical Laboratory, Tianjin Medical University, Tianjin 300203, China; 2. Department of Clinical Laboratory, The Second Hospital, Tianjin Medical University, Tianjin 300203, China)

Abstract Objective: To evaluate the consistency of the clinical diagnostic value and diagnostic efficacy of serum thyroid function test items among various imported pharmaceutical brands, including Roche Cobas E602, Siemens Centaur XP and Beckman DXI800. Methods: We selected 63 cases diagnosed hyperthyroidism, 66 cases diagnosed hypothyroidism, as well as 70 cases not diagnosed either of the aforementioned diseases as the control group. We collected untreated serum samples from the three groups and tested five thyroid functions (TSH, FT4, FT3, T4, T3). The diagnostic value and efficacy of the treatment was evaluated based on the Receiver Operating Characteristic curve (ROC curve) of multi-factor combined diagnosis and the McNemar scheme highlighted in WS/T505-2017 Guide of Qualitative Determination Performance Evaluation. Results: Based on the analysis of the ROC curve, five projects cooperative detection with three brands, the AUC of patients with hyperthyroidism was between 0.98 and 0.99, and the AUC of patients with hypothyroidism was between 0.81 and 0.84; there was no statistical difference among the three brands (all P>0.05). Based on McNemar scheme, considering the established critical point (upper and lower bound of the reference range) of each brand, the diagnostic efficacy of the three brands for the hyperthyroid patients was virtually the same, except TSH, Siemens demonstrated better clinical sensitivity and lower clinical specificity. The three brands showed different advantages when used to treat hypothyroidism patients. Furthermore, the clinical sensitivity and specificity shown on FT₃ and FT₄ were relatively low (13.70%-50.00%) for all three brands, regardless of whether they were used against hyperthyroidism or hypothyroidism. Conclusion: In terms of the testing of serum thyroid function, Cobas E602, Centaur XP and DXI800 all have high diagnostic value and are consistent for detecting hyperthyroidism. Hence, we would suggest clinical laboratory organizations that utilize these three brands for hyperthyroidism testing to consider mutual acceptance of testing results. In addition, although the three brands all have medium-level diagnostic value and are consistent for detecting hypothyroidism, since the clinical sensitivity on FT₃ and FT₄ is low for all three, we believe that diagnosis efficacy still needs to be improved.

Key words thyroid function test; McNemar scheme; ROC curve; clinical diagnostic value

甲状腺疾病是临床常见的内分泌疾病,近年来甲状腺功能亢进症(甲亢)及甲状腺功能减退症(甲减)的发病率逐渐增加,严重影响患者日常生活及工作[1-2]。《中国甲状腺疾病诊治指南》中对于甲亢及甲减的诊断必要条件均包括血清甲状腺功能检测。

临床中常见不同品牌间血清甲状腺功能检测结果存在差异,而当前研究往往重视检测结果数值的一致性^[3],或是对进口仪器进行性能验证^[4],但不同品牌检测系统由于其参考范围相差较大,理论上不宜开展对于血清甲状腺功能检测数值的可比性及一致性分析^[5]。本研究依据多因素联合诊断 ROC 曲线及 WS/T505-2017《定性测定性能评价指南》提供的 McNemar 方案^[6],评估不同进口品牌间 5 项血清甲状腺功能检测项目临床诊断价值及诊断效能是否一致,探讨在不同检测机构间开展结果互信是否具有可行性。

1 资料与方法

1.1 临床资料 收集我院 2018 年 3 月—2019 年 4 月,初诊诊断为甲亢患者 63 例,甲减患者 66 例,患者均符合典型的临床诊断特征,甲状腺功能指标均符合甲亢或甲减的诊断标准[□],同时均未进入治疗期,选择与甲亢组及甲减组年龄、性别相匹配的健康体检人群 70 名作为对照组,共计收集标本 199个,每个标本分装 3 份—80℃保存。

1.2 试验方案 分别选择血清甲功检测市场占有率较大的 3 个进口品牌及仪器:罗氏 Cobas E602(电化学发光法)、西门子 Centaur XP(吖啶酯直接化学发光法)及贝克曼库尔特 DXI800(碱性磷酸酶-金刚烷酶促化学发光法),分别检测分装样本中促甲状腺激素(thyroid-stimulating hormone, TSH)、FT₄、FT₃、T₄及T₃水平,保障各品牌试验日室内质控结果在控,同时试验仪器系统封闭,可溯源,仪器依据实验室 SOP 文件定期执行周期性校准及保养维护。所有样本在每个品牌仪器上重复检测 2 次,2 d 内检测完毕,记录均值。

1.3 多因素联合诊断 ROC 曲线评价诊断价值 分别将 Cobas E602、Centaur XP 及 DXI800 检测的血清 TSH、 FT_4 、 FT_3 、 T_4 、 T_3 均值记录,对每个品牌的 5个检测结果采用 *Logistic* 回归计算合成概率,将临床

Tab 1 Summary of the reference range for three brands

表 1 3 个品牌参考范围汇总

诊断甲亢或甲减作为"金标准",绘制 ROC 曲线,比较各品牌曲线下面积(AUC)。同时分析以上形成的 AUC 主要贡献来源,本研究将每个品牌的 5 个检测项目分别组合或单独成组,采用 Logistic 回归计算相应合成概率,绘制 ROC 曲线,比较不同组合项目 AUC 与 5 个项目联合 AUC 是否有统计学差异,进而判断贡献 AUC 的主要检测项目。

1.4 McNemar 方案评价诊断效能 ROC 曲线考虑 了所有可能的诊断临界点的影响,而结合各品牌既 定的临界点(参考范围上下限),采用 McNemar 方案 评价诊断效能。该方案常用来判断两种方法的临床 敏感性和临床特异性之间是否有统计学差异。这种 统计学检验没有假定一种方法优于另一种方法,而 是认为两种方法均可能引起诊断错误,将两方法同 时与诊断准确度标准比较,形成配对数据,进而计 算两方法的临床敏感性之差及临床特异性之差的 可信区间。本研究依据各品牌临界点(表 1),结合甲 亢、甲减数据变化特点,分别针对甲亢、甲减判断数 据组阴阳性,将临床诊断作为金标准,3个品牌两两 比较计算临床敏感性及临床特异性。若两系统间临 床敏感性之差及临床特异性之差的可信区间包括 0,认为两品牌临床敏感性或临床特异性一致,诊断 效能相当。

1.5 统计学方法 应用 SPSS20.0 软件中二项 Logistic 回归计算合成概率,MedCalc 软件绘制 ROC 曲线并比较不同品牌间 AUC 大小。应用 Excel 软件计算临床敏感性、临床特异性以及各自差值的置信区间。P<0.05 为差异有统计学意义。

2 结果

2.1 多因素联合诊断 ROC 曲线分析

2.1.1 3 个品牌多项目联合检测对甲亢疾病的诊断价值—致性评价 5 个项目(TSH、 FT_4 、 FT_3 、 T_4 、 T_3)联合检测ROC 曲线见图 1,不同品牌 AUC:Cobas=0.99、Centaur=0.98、DXI=0.99,三品牌两两比较 AUC 之间 差异无统计学意义(均 P>0.05)。

为比较个别项目对 AUC 的主要贡献,将每个品牌的 5 个项目联合、3 个项目(TSH、FT₄、FT₃)联合及单独检测 TSH 绘制 ROC 曲线,具体 AUC 结果及 AUC 差异比较见表 2,可得 3 个品牌单独检测TSH

Tab 1 Summary of the reference range for three brands									
品牌	$TT_3/(nmol/L)$	$TT_4/(nmol/L)$	FT ₃ /(pmol/L)	FT ₄ /(pmol/L)	$TSH/(\mu IU/mL)$				
CentaurXP	0.92~2.79	58.94~140.60	3.50~6.50	11.50~22.70	0.55~4.78				
CobasE602	1.30~3.10	66.00~181.00	3.10~6.80	12.00~22.00	0.27~4.20				
DXI800	1.34~2.73	78.38~157.40	3.80~6.00	8.50~15.41	0.34~5.60				

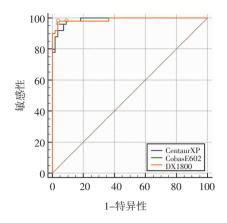


图 1 3 个品牌 5 个项目甲亢联合检测多因素 ROC 图

Fig 1 The ROC curve of multi-factor combined diagnosis for five projects cooperative detection with three brands in hyperthyroidism

的 AUC 均与 5 项或 3 项联合检测 AUC 无差异(均 P>0.05),在初诊患者中对甲亢的诊断价值 3 个品牌基本一致,同时 TSH 的检测贡献了绝大部分 AUC。2.1.2 3 个品牌多项目联合检测对甲减疾病的诊断价值一致性评价 5 个项目(TSH、FT₄、FT₃、T₄、T₃)联合检测 ROC 曲线见图 2,不同品牌 AUC 为 Cobas =0.84、Centaur=0.832、DXI=0.81,3 个品牌两两比较 AUC 之间差异无统计学意义(均 P>0.05)

同样,为比较个别项目对 AUC 的主要贡献,将 每个品牌的 5 个项目联合、3 个项目 1(TSH、FT₄、 FT₃)联合、3 个项目 2(TSH、T₄、T₃)联合、单独检测 TSH、单独检测 FT₄ 及单独检测 FT₃ 绘制 ROC 曲线,

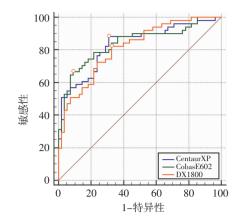


图 2 3 个品牌 5 个项目甲减联合检测多因素 ROC 图

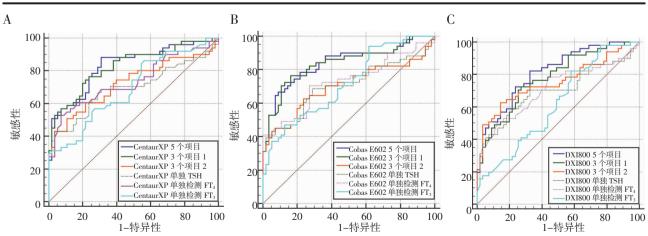
Fig 2 The ROC curve of multi-factor combined diagnosis for five projects cooperative detection with three brands in hypothyroidism

具体见图 3A-C, 可得 Centaur XP 5 个项目(AUC=0.83)与 3 个项目 1(AUC=0.82)的 AUC 无差异,与其他组合及单项比较(AUC 在 0.69~0.73 之间)差异明显(均 P>0.05);CobasE602 5 个项目(AUC=0.84)与 3 个项目 1(AUC=0.83)的 AUC 无差异,与其他组合及单项比较(AUC 在 0.70~0.72 之间)差异明显(均 P>0.05);DXI800 5 个项目(AUC=0.81)与 3 个项目 1(AUC=0.77)的 AUC 无差异,与其他组合及单项比较(AUC 在 0.61~0.72 之间)差异明显(均 P>0.05)。可得在初诊患者中对甲减的诊断 3 个品牌价值中等并基本一致,同时 TSH、 FT_4 、 FT_3 的检测贡献了绝大部分 AUC。

表 2 甲亢患者每个品牌不同项目联合计算得 AUC 及 AUC 差异

Tab 2 AUC calculated by different cooperative detected projects of every brand in hyperthyroidism and its difference

品牌		不同项目联合	7	n	
	5 个项目联合	3 个项目联合	单独 TSH	Z	Ρ
CentaurXP	0.98	0.99	0.98	0.10~0.27	>0.05
CobasE602	0.99	0.98	0.98	0.11~0.31	>0.05
DXI800	0.99	0.99	0.99	0.10~0.21	>0.05



注: A: Centaur XP; B: Cobas E602; C: DXI800)

图 3 甲减患者每个品牌不同项目联合绘制 ROC 曲线

Fig~3~~The~ROC~curve~drawn~by~different~cooperative~detected~projects~of~every~brand~in~hypothyroidism~

2.2 McNemar 方案评价两两品牌间临床敏感性及临床特异性 针对不同品牌参考区间,依据 2.1 结果,选取影响 AUC 较大的 TSH、FT4、FT3 项目,判断甲亢及甲减患者 3 项目检测结果阴阳性,3 个品牌两两配对,比较配对两品牌之间临床敏感性及临床特异性,见表 3。对于甲亢患者:TSH 项目Centaur 的临床敏感性优于 Cobas 及 DXI, Cobas 及 DXI 的临床

特异性优于 Centaur,其他所有比对临床敏感性及特异性均一致。对于甲减患者:FT₃项目,Centaur 的临床敏感性优于 Cobas 与 DXI,临床特异性基本一致;FT₄项目,临床敏感性 DXI 优于 Cobas,Cobas 优于 Centaur,临床特异性 DXI 明显低于 Cobas 与 Centaur;TSH 项目,临床敏感性 Cobas 优于Centaur,Centaur 优于 DXI,临床特异性与之呈相反关系。

表 3 McNemar 方案评价两品牌间临床敏感性及临床特异性

Tab 3 Evaluate the clinical sensitivity and specificity between the two brands that based on McNemar scheme

疾病	番目	证 (人	比较 1/%		比较 2/%			比较 3/%			
	项目	评价指标	Centaur	Cobas	差值置信区间	Centaur	DXI	差值置信区间	Cobas	DXI	差值置信区间
甲亢	FT_3	临床敏感性	40.00	40.00	-6.69~6.69	40.00	40.00	-6.69~6.69	40.00	40.00	-6.69~6.69
		临床特异性	100.00	100.00	-	100.00	94.50	-	100.00	94.53	_
	FT_4	临床敏感性	42.10	50.00	-16.03~0.36	42.00	48.10	-16.42~4.75	50.00	48.00	-5.50~9.49
		临床特异性	100.00	96.36	-	100.00	94.52	-	100.00	94.55	-
	TSH	临床敏感性	98.00	86.30	1.79~24.05*	98.10	90.00	1.26~19.16*	86.00	90.20	-12.93~3.86
		临床特异性	92.60	100.00	-	92.72	98.17	-15.24~2.58	100.00	98.16	-
甲减	FT_3	临床敏感性	27.45	13.71	3.44~24.71*	27.46	15.69	1.94~25.26*	13.72	15.69	-12.39~8.25
		临床特异性	100.00	100.00	-	100.00	98.18	_	100.00	98.18	-
	FT_4	临床敏感性	23.57	39.20	-26.20~-4.90*	23.53	50.97	-39.24~-14.15*	39.20	50.98	-20.82~-2.16*
		临床特异性	100.00	100.00	-	100.00	85.45	_	100.00	84.45	_
	TSH	临床敏感性	60.74	70.58	-18.72~-0.77*	60.75	47.01	3.52~23.23*	70.56	47.04	10.94~34.76*
		临床特异性	74.52	61.81	3.18~22.16*	74.56	83.64	-18.38~-0.49*	61.82	83.63	-33.18~-10.18*

注:* 差值置信区间不包括 0, 差异有统计学意义: 甲亢: 甲状腺功能亢进症: 甲减: 甲状腺功能减退症

3 讨论

甲状腺疾病是临床常见病,作为内分泌领域第二大疾病,其发病率逐渐增加,以甲亢和甲减最为常见。甲亢是由于甲状腺腺体本身功能亢进,合成和分泌甲状腺激素增加所致的甲状腺毒症,其可加速机体新陈代谢,增加机体能量消耗;甲减是由于甲状腺激素合成和分泌减少或组织利用不足导致的全身代谢减低综合征。无论是何种类型疾病,均对患者身体健康构成一定危害,因此甲亢及甲减的二级预防应引起足够重视,早发现、早诊断、早治疗对改善患者生存质量具有重要意义[8-9]。

临床医生日常诊治过程中,常发现不同进口品牌仪器间的血清甲状腺功能检测结果数值一致性较差,并且不同检测项目的组合在疾病诊断中的异常表现也不尽相同,导致品牌间的结果互认较难开展,造成大量重复检测。

本研究利用多因素联合诊断 ROC 曲线分析市场占有率较大的 3 个品牌即罗氏、西门子及贝克曼之间在甲亢及甲减项目中的诊断价值。所得结论为:对于甲亢的诊断,3 个品牌 5 个项目联合检测的

临床诊断价值均较高并一致(AUC 在 0.98~0.99 之间),同时 TSH 的单独检测结果贡献了绝大部分曲线下面积(AUC 在 0.98~0.99 之间),即 3 个品牌同时检测 TSH、FT4、FT3、T4、T3 与单独检测 TSH 的AUC 无统计学差异,故在甲亢的初诊诊断中,TSH的诊断价值尤为重要,同时 3 个品牌能够做到该项目的诊断价值一致;对于甲减的诊断,3 个品牌 5 个项目联合检测的临床诊断价值均中等并一致(AUC 在 0.81~0.84 之间),同时经过不同项目组合筛选,3 个品牌 TSH、FT4、FT3 联合检测与 5 个项目联合检测相比,AUC 无差异,故在甲减的初诊诊断中,TSH、FT4、FT3 的诊断价值尤为重要,同时 3 个品牌能够做到该 3 个项目的诊断价值一致,但其对于甲减的诊断价值不及甲亢。

本研究同时依据各品牌的参考范围判断各数据阴阳性,以临床诊断为金标准比较3个品牌的诊断效能。所得结论为:对于甲亢,TSH项目临床敏感性西门子>罗氏=贝克曼,临床特异性罗氏=贝克曼>西门子,其他临床敏感性和特异性均一致,故结合参考范围的设定,3个品牌的TSH、FT4、FT3,临床诊

断效能基本一致,其对临界点的设定能够满足对甲 亢的诊断并诊断效能基本一致,西门子在获得更好 临床敏感性的同时放弃了一部分临床特异性。对于 甲减,FT3临床敏感性西门子>罗氏=贝克曼,均较低 (13.70%~27.41%), 临床特异性基本一致;FT4临床 敏感性贝克曼>罗氏>西门子, 临床特异性西门子= 罗氏>贝克曼:TSH 临床敏感性罗氏>西门子>贝克 曼,临床特异性贝克曼>西门子>罗氏。故3个品牌 对临界点的设定,在甲减的诊断中互有优势,西门 子FT,项目临床敏感性优于其他两个品牌,但FT。临 床敏感性最低,TSH 临床敏感性中等水平;罗氏 TSH 临床敏感性优于其他两个品牌,但FT4临床敏感性 中等水平,FT,临床敏感性与贝克曼持平但稍差于 西门子;贝克曼 FT4 临床敏感性优于其他两个品牌, 但 TSH 临床敏感性最低,可见,单次血清 TSH 检测 值作为亚甲减的诊断依据存有技术不确定性,这与 徐清芳等[10]、王婷婷等[11]的研究所得结论类似。通过 表 3 可得出无论是甲亢或甲减疾病,以品牌既定临 界值(参考范围上下限)辅助诊断疾病,FT,及FT。的 临床敏感性均较低,临床诊断效能有待提高。

综上所述,罗氏 Cobas E602、西门子 Centaur XP 及贝克曼 DXI800 检测血清甲状腺功能项目,对于甲亢的诊断价值较高并基本一致,对于甲减的诊断价值中等并一致;同时以品牌既定临界点辅助诊断甲减,从临床敏感性及临床特异性的角度来看,3个品牌互有优势;但在 FT3 及 FT4 的检测上,无论是甲亢或甲减疾病,3 个品牌临床敏感性均较低,临床诊断效能有待提高。基于上述研究,在甲亢的临床诊断方面,可建议使用该 3 种品牌的检验机构开展检验结果互信,以减少重复检验,节约医疗资源。

参考文献:

- [1] 黄杰,于鹏飞,杨志平.促甲状腺激素和甲状腺自身抗体在甲状腺 疾病诊断中的临床应用[J].中国地方病防治杂志,2015,13(1):33
- [2] 连小兰.促甲状腺激素相关性甲状腺疾病[J]. 中国实用内科杂志, 2014,34(4):325
- [3] 林明, 靳平燕, 刘丽芳. 不同化学发光免疫分析设备甲功五项检测结果比对分析[J]. 标记免疫分析及临床, 2019, 26(9): 1599
- [4] 李悦,陈亚琼,覃俊龙,等.雅培 i2000SR 全自动化学发光分析仪 检测甲状腺功能五项的性能验证[J]. 检验医学与临床,2019, 16(22):3243
- [5] Clinical and Laboratory Standards Institute. Measurement procedure comparison and bias estimation using patient samples; approved guideline – third edition. CLSI document EP9-A3[S]. Clinical and Laboratory Standards Institute, 940 West Valey Road, Suite 1400, Wayne, Pennsylvania 19087–1898 USA, 2005
- [6] 中华人民共和国国家卫生和计划生育委员会. WS/T505-2017 定性测定性能评价指南[S]. 北京: 中华人民共和国国家卫生和计划生育委员会,2017
- [7] Ross D S, Burch H B, Cooper D S, et al. 2016 American thyroid association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis[J]. Thyroid, 2016, 26(10):1343
- [8] Biswas D, Dutta D, Maisnam I, et al. Occurrence of osteoporosis& factors determining bone mineral loss in young adults with Graves' disease[J].Indian J Med Res, 2015, 141(3):322
- [9] Yoshihara A, Yoshimura N J, Mukasa K, et al. The characteristics of osteoporotic patients in Graves' disease patients newly diagnosed after menopause: a prospective observational study [J]. Endocr J, 2016, 63(12):1113
- [10] 徐清芳,张美华,姚文杰,等.不同化学发光免疫分析系统检测 亚临床甲状腺功能减退症血清促甲状腺素的对比研究[J].实用 医技杂志,2019,26(9):1140
- [11] 王婷婷,单凤玲,陆汉魁.不同化学发光免疫分析系统检测亚临 床甲状腺功能减退症血清促甲状腺激素的结果对比[J].中华核 医学与分子影像杂志,2017,37(6):342

(2020-02-01 收稿)

欢迎投稿 欢迎订阅