

MR 扩散加权成像对乳腺病变良恶性鉴别的 Meta 分析报告

陈欣 张毅力 张秋娟 郭佑民 赵暹 白芝兰

【摘要】 目的 采用 Meta 分析方法综合定量评价 MR 扩散加权成像(DWI)对乳腺良恶性病变的鉴别诊断价值。方法 检索 Cochrane 图书馆、Pubmed、Ovid 循证医学数据库、Elsevier 和 Springer 数据库及中国期刊网 1997 至 2007 年公开发表的中英文文献,按照 Cochrane 协作网推荐的诊断试验纳入标准筛选文献,提取纳入研究的特征信息。采用 Q 检验和 Meta 分析对纳入文献行统计学处理。结果 共纳入文献 6 篇,研究病灶共计 479 个。DWI 诊断乳腺病变良恶性的敏感度范围为 64.0% ~ 92.8%,特异度范围为 45.8% ~ 96.7%。纳入文献具同质性,按照固定效应模型计算出的汇总加权敏感度和特异度及 95% 可信区间分别为 86% (76% ~ 92%) 和 80% (61% ~ 91%);汇总受试者工作特征曲线(SROC)下面积为 92.19%。敏感度分析显示纳入文献稳定性好。结论 DWI 的表观扩散系数(ADC)值测量有助于乳腺病变的良恶性鉴别,但缺乏统一的扩散敏感因子(b)值和诊断阈值,其临床应用价值仍有待于前瞻性研究进一步证实。

【关键词】 乳腺疾病; 磁共振成像; 扩散; Meta 分析

Analysis of diffusion-weighted MR in the differential diagnosis of benign and malignant breast lesions with Meta-analysis CHEN Xin*, ZHANG Yi-li, ZHANG Qiu-juan, GUO You-min, ZHAO Xian, BAI Zhi-lan. * Imaging Center, Second Hospital, Medical College of Xi an Jiaotong University, Xi an 710004, China

Corresponding author: GUO You-min, Department of Radiology, Chaoyang Hospital, Capital University of Medical Sciences, Beijing 100020, China, Email: cjr. guoyoumin@vip. 163. com

【Abstract】 Objective To evaluate the value of diffusion-weighted imaging (DWI) in the differential diagnosis of benign and malignant breast lesions with Meta-analysis. Methods English and Chinese articles published from 1997 to 2007 were searched in Cochrane library, Pubmed, Ovid, Elsevier, Springer databases and Chinese National Knowledge Infrastructure (CNKI), and selected according to the criteria for diagnostic research recommended by Cochrane Method Group. The characteristics of selected articles including the study background, design, results and diagnostic indicators were analyzed. Statistical analysis was performed with Meta-test version 0.9. Results Six articles were included with a total of 479 lesions. The sensitivity and specificity of DWI for breast lesions ranged from 64.0% to 92.8% and from 45.8% to 96.7% respectively. All groups had homogeneity, and a fixation effects regression was developed. The pooled weighted sensitivity and specificity with 95% CI and Az was 86% (76% —92%), 80% (61% —91%) and 92.19% respectively. Sensitivity analysis demonstrated no remarkable change on heterogeneity. Conclusion The value of ADC in DWI can help to differentiate the benign breast lesions from the malignant with good sensitivity and specificity. The further study is needed to assess the clinical application of the b-value and diagnostic threshold.

【Key words】 Breast disease; Magnetic resonance imaging; Diffusion; Meta-analysis

乳腺癌是全世界妇女,也是中国妇女常见的恶

性肿瘤之一^[1],早期、准确地判断病变性质对治疗和预后均有重要的意义。MR 扩散加权成像(diffusion-weighted imaging, DWI)已广泛应用于急性脑梗死、脑卒中的研究,初步应用于肝脏、胰腺、子宫和乳腺等病变的研究。笔者采用 Meta 分析方法分析国内外 DWI 鉴别乳腺良恶性病变的研究结果,以综合定量评价 DWI 对乳腺良恶性病变的鉴别诊

基金项目:国家自然科学基金资助项目(60501006、30701008);
陕西省科学技术研究发展计划项目[2007K09-06(5)]

作者单位:710004 西安交通大学医学院第二医院影像中心(陈欣、张秋娟、赵暹、白芝兰);西安交通大学医学院第一医院影像中心(张毅力);首都医科大学附属朝阳医院放射科(郭佑民)

通信作者:郭佑民,100020,Email: cjr. guoyoumin@vip. 163. com

断价值及影响因素。

资料与方法

一、文献检索与筛选

1. 资料来源: 检索的数据库包括 Cochrane 图书馆、Pubmed、Ovid 循证医学数据库、Elsevier 和 Springer 数据库及中国期刊网。发表年限为 1997 年 1 月至 2007 年 3 月。本研究不包括尚未公开发表的文章。英文检索词为“DWI OR diffusion-weighted imaging AND breast OR mammary”; 语种限定为英文和中文, 研究对象限定在“人类”, 并将发表类型为“综述”、“信件”、“评论”、“社论”和“病例报告”的文章排除。中文检索词为“弥散加权成像”、“扩散加权成像”和“乳腺”。为尽量避免漏查文献, 采用网上检索及手工检索相结合的方法, 对所检索的文献中提供的参考文献进行二次检索。由 2 名评价员独立进行检索, 如遇不一致经讨论确定。评价员为医学影像学专业博士研究生, 且已系统地阅读乳腺病变各种影像学诊断的相关文献。

2. 纳入标准: 文献筛选根据 Cochrane 协作网筛选与诊断试验方法组中关于诊断试验性研究的纳入标准进行^[2], 具体纳入标准包括: (1) 研究目的为评价或探讨 DWI 鉴别乳腺病变良恶性的诊断价值; (2) 研究对象包括乳腺病变所有常见的病理类型; (3) 研究对象 30 例, 且良恶性病变的例数分别 10 例; (4) A 级或 B 级的文献, 包括前瞻性或回顾性研究; (5) 有病理学结果作为诊断的金标准; (6) 能直接或间接获得 DWI 诊断乳腺病变良恶性的真、假阳性值及真、假阴性值。

二、资料内容提取

1. 研究背景和设计信息的提取: 包括研究作者、发表时间、研究国家; 研究对象的数量、平均年龄、年龄范围; 研究的连续性和类型(前瞻性或回顾性)。

2. DWI 结果信息的提取: 包括扩散敏感因子(b)值, 良性和恶性病变的平均表观扩散系数(apparent diffusion coefficient, ADC)值, 正常乳腺的平均 ADC 值, 区别良恶性病变的 ADC 阈值。

3. 诊断参数信息的提取: 包括真阳性值(true positive, TP)、假阳性值(false positive, FP)、真阴性值(true negative, TN)、假阴性值(false negative, FN)、敏感度(sensitivity, Sen)、特异度(specificity, Spe)、诊断准确度(accuracy, Acc)、阳性预测值(positive predictive value, PPV)、阴性预测值(negative predictive value, NPV)、阳性似然比

(positive likelihood ratio, PLR) 和阴性似然比(negative likelihood ratio, NLR)。

三、数据分析

1. 异质性检验: 采用 Q 检验, Q 服从自由度为 $k - 1$ 的 χ^2 分布, Q 值越大, 其对应 P 值越小。若 $Q > \chi^2_{(1-\alpha)}$, 则 $P < \alpha$, 表明纳入研究间存在异质性, 应选择随机效应模型进行 Meta 分析; 反之, 若 $Q < \chi^2_{(1-\alpha)}$, 则 $P > \alpha$, 表明纳入研究具有同质性, 应选择固定效应模型进行 Meta 分析。

2. Meta 分析: 按照相应的效应模型, 首先将各个研究的敏感度和特异度进行 Logit 变换, 再按照权重大小进行汇总, 最后进行反 Logit 变换得出加权汇总敏感度和特异度及相应的 95% 可信区间, 并绘制森林图。

3. 建立汇总受试者工作特征曲线(SROC): 绘制 SROC 曲线, 并计算其曲线下面积(Az), 曲线越接近坐标轴左上角, Az 则越接近 100%, 说明该检查的诊断价值越高。

4. 敏感度分析: 将结果可疑的文献排除后, 重新进行 Meta 分析, 所得结果与未排除前的结果进行比较, 若敏感度分析未从实质上改变结果, 即汇总敏感度、特异度和纳入文献的异质性未发生实质性改变, 说明纳入文献的稳定性好, 结果较为可信; 反之, 提示有潜在的重要因素干预结果, 在解释结果和得出结论时应非常慎重, 须明确争议的来源^[3]。

结 果

一、文献检索与质量分析

共检索到文献 21 篇, 其中英文文献 11 篇, 中文文献 10 篇。按照纳入标准最终选定英文文献 2 篇, 排除 9 篇, 其中与研究目的不符者 5 篇, 研究目的仅为乳腺癌和纤维腺瘤的鉴别诊断者 1 篇, 研究对象仅为乳腺癌者 1 篇, 研究对象不足 20 例者 1 篇, 无法得到诊断信息参数者 1 篇。选定中文文献 4 篇, 排除 6 篇, 其中综述 2 篇, 病例数不足者 1 篇, 研究目的不符者 3 篇。全部 6 篇纳入文献中, 前瞻性研究 1 篇, 回顾性研究 5 篇。

二、数据提取

研究病灶共计 479 个, 良性病灶的平均 ADC 值范围为 $1.49 \sim 1.67 \times 10^{-3} \text{ mm}^2/\text{s}$, 恶性病灶的平均 ADC 值的范围为 $0.91 \sim 1.22 \times 10^{-3} \text{ mm}^2/\text{s}$, 良恶性鉴别的 ADC 阈值范围为 $1.01 \sim 1.6 \times 10^{-3} \text{ mm}^2/\text{s}$, DWI 诊断乳腺病变良恶性的敏感度范围为 64.0% ~ 92.8%, 特异度范围为 45.8% ~ 96.7%。

纳入文献的提取信息见表 1 ~3。

三、数据分析

1. 异质性检验: $Q = 5.66$, 自由度 () = 5, $P > 0.250$, 因纳入研究数目较少, 为增大检验效能, 选择检验水准 $\alpha = 0.10$, 接受同质性假设, 采用固定效应模型进行加权定量合并。

2. Meta 分析: 汇总加权敏感度和特异度及 95% 可信区间分别为 $0.86 (0.76 \sim 0.92)$ 和 $0.80 (0.61 \sim 0.91)$ (图 1)。

3. 建立 SROC 曲线 (图 2): SROC 的 A_z 为 92.19%。

4. 敏感度分析: 将诊断敏感度和特异度与其他研究结果差异较大的文献^[4,9]逐一排除后, 文献的异质性、汇总敏感度和特异度见表 4。

讨 论

DWI 有助于乳腺病变鉴别诊断^[10], 但文献报道不多, 病例数较少, 大多数为回顾性研究, 国内外均未见有关 DWI 鉴别乳腺病变方面的系统回顾性文献。笔者采用系统评价的方法和 Meta 分析软件, 按照循证医学的要求将其中高质量的文献纳入研究, 以增加结果的可信性, 在计算出汇总加权敏感度和特异度的同时, 还绘制了 SROC, 力求对 DWI 鉴别乳腺良恶性病变的诊断价值进行更加全面的量化分析, 为后续研究和临床决策提供指导。

本研究纳入文献中, 5 篇 $b_{高} = 1$ 篇 $b_{高} = 800 \text{ s/mm}^2$ ^[7], 二者的诊断敏感度和特异度相当。虽然不同 b 值下测量的 ADC 值差异有统计学意

表 1 纳入文献研究背景和设计信息

研究作者	发表时间	资料来源	患者平均年龄(岁)	年龄范围(岁)	病灶数(个)	连续性	研究类型
Woodhams 等 ^[4]	2005	日本	53	14 ~88	191	未提	回顾性
Rubesova 等 ^[5]	2006	比利时	52	25 ~74	87	是	前瞻性
郭勇等 ^[6]	2001	中国	45	34 ~70	46	未提	回顾性
罗建东等 ^[7]	2007	中国	43	24 ~65	60	是	回顾性
李洁等 ^[8]	2005	中国	45	24 ~68	41	是	回顾性
赵斌等 ^[9]	2005	中国	43	21 ~72	54	是	回顾性

表 2 纳入文献 ADC 值测量结果信息($\times 10^{-3} \text{ mm}^2/\text{s}$, 珉 \pm s)

研究作者	扩散敏感因子(s/mm^2)	恶性病变	良性病变	正常乳腺	ADC 阈值
Woodhams 等 ^[4]	1000	1.22 \pm 0.31	1.67 \pm 0.54	2.09 \pm 0.27	1.600
Rubesova 等 ^[5]	1000	0.95 \pm 0.02	1.51 \pm 0.07	—	1.150
郭勇等 ^[6]	1000	0.96 \pm 0.20	1.59 \pm 0.32	1.76 \pm 0.40	1.297
罗建东等 ^[7]	800	0.87 \pm 0.23	1.59 \pm 0.26	1.98 \pm 0.31	1.220
李洁等 ^[8]	1000	1.21 \pm 0.26	1.49 \pm 0.43	—	1.420
赵斌等 ^[9]	1000	0.91 \pm 0.25	1.58 \pm 0.22	1.78 \pm 0.51	1.010

注: ADC 为表观扩散系数, —为无数据

表 3 纳入文献诊断试验参数信息

研究作者	TP(个)	FP(个)	TN(个)	FN(个)	Sen(%)	Spe(%)	Acc(%)	PPV(%)	NPV(%)	PLR	NLR
Woodhams 等 ^[4]	155	12	11	13	92.8	45.8	86.9	92.3	47.8	1.71	0.16
Rubesova 等 ^[5]	53	3	19	12	81.5	86.4	74.2	94.6	61.3	6.0	0.21
郭勇等 ^[6]	24	3	17	2	92.3	85.0	89.1	88.9	89.5	6.2	0.09
罗建东等 ^[7]	24	4	29	3	88.9	87.9	88.3	85.7	90.6	7.4	0.13
李洁等 ^[8]	25	5	8	3	89.3	61.5	80.5	83.3	72.7	2.3	0.17
赵斌等 ^[9]	15	1	29	9	64.0	96.7	81.5	93.8	76.3	19.4	0.37

注: TP 为真阳性值、FP 为假阳性值、TN 为真阴性值、FN 为假阴性值, Sen 为敏感度, Spe 为特异度, Acc 为诊断准确度, PPV 为阳性预测值, NPV 为阴性预测值, PLR 为阳性似然比, NLR 为阴性似然比

表 4 排除结果异常的研究后 Q 值、P 值及异质性分析和汇总敏感度及特异度

纳入文献	Q 值	P 值	异质性分析	敏感度(%)	95% 可信区间(%)	特异度(%)	95% 可信区间(%)
全部文献	5.66	>0.25	同质性	86	76 ~92	80	61 ~91
排除 Woodhams 等 ^[4]	2.38	>0.50	同质性	83	72 ~91	85	72 ~92
再排除赵斌等 ^[9]	2.27	>0.50	同质性	86	79 ~91	82	69 ~90

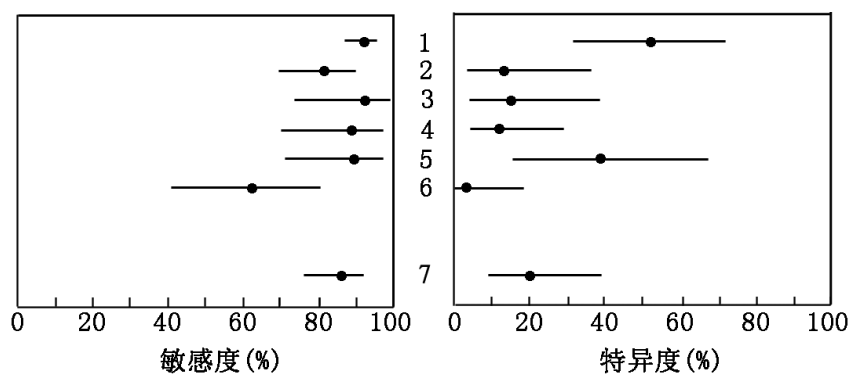


图 1 DWI 诊断乳腺良恶性病变的敏感度和特异度的森林图。1~6 所对应的圆点分别表示 6 个纳入研究的敏感度和特异度, 其两侧的直线为相应的 95% 可信区间。直线分布越窄, 说明该研究的精度越大, 在加权时给予的权重越大。7 表示汇总敏感度和特异度及相应的 95% 可信区间

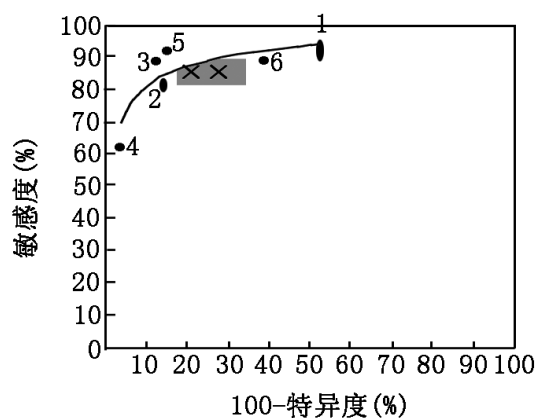


图 2 纳入研究的 SROC 曲线。图内 1~6 黑斑为根据各个研究的敏感度和特异度绘制的散点, 其大小表示按精度给予的权重。××及所在的灰色区域为汇总加权敏感度和特异度及 95% 可信区间

义^[9], 但对鉴别良恶性病变的诊断效能有否影响尚未见报道, 且本组研究中 $b_{高} = 800s/mm^2$ 的文献数量过少无法进一步分层分析。笔者建议采用多个较大的 $b_{高}$ 值分别对乳腺病灶进行研究, 比较之间的差异和诊断效能, 希望从中能够确定一个最佳的 $b_{高}$ 值。

关于 DWI 鉴别乳腺病变良恶性的敏感度和特异度, 各家报道差异较大, 分析原因, 可能是诊断阈值的选择方法不同和研究对象选择的偏倚。纳入文献选取的 ADC 阈值最低为 $1.01 \times 10^{-3} mm^2/s$ ^[9], 最高则为 $1.6 \times 10^{-3} mm^2/s$ ^[4]。其中赵斌等^[9]将恶性肿瘤的 95% CI 上界作为界值, 敏感度偏低 (64.0%); Woodhams 等^[4]如何选择诊断阈值文献中没有说明, 其特异度偏低 (45.8%); 郭勇等^[6]以 $\pm s$ 为标准, 其余文献采用 ROC 曲线法获得 ADC 诊断阈值, 其敏感度相似^[2,7-8], 而李洁等^[8]的研究对象多数为临床检查或乳腺 X 线检查为可疑恶性或恶性的病例, 病理证实为良性的病变多数为临床、X 线、超声诊断困难的病变, 所以特异度偏低 (61.5%)。Meta 分析结果显示, 汇总敏感度和特异度及 95% CI 分别为 86% 和 80%。敏感度分析在

逐一排除结果差异最大的研究后, 文献的异质性无实质性改变, 汇总敏感度和特异度差异无统计学意义, 说明本研究评价出的结果可信。同时, 为了更准确地反映敏感度和特异度间的线性关系, 根据各研究的比值比 (OR) 的权重绘制 ROC 曲线, 即 SROC 曲线。SROC 曲线不受诊断阈值变化的影响, 可以通过图形和面积直接反映诊断性试验的效能^[11]。SROC 的 Az 代表诊断试验的准确性, 本研究得出的 SROC 图形符合判定为诊断价值较高的检查方法所应得到的图形, Az 表明 DWI 诊断的准确度为 92%, 进一步肯定了 DWI 鉴别乳腺良恶性病变的价值。

目前尚无统一的 ADC 诊断阈值标准, 汇总分析发现国人资料 ADC 阈值在 $1.2 \times 10^{-3} mm^2/s$ 左右的诊断效能最高, 敏感度和特异度均较高, 可以尝试以此作为阈值标准对乳腺病变性质的鉴别作前瞻性研究, 进一步评价其临床应用价值。

本研究采用系统评价的方法和 Meta 分析软件, 克服了单个研究在病例来源、研究设计、统计分析等方面差异较大、研究样本偏小的缺点, 增加了结果的可信度, 但由于相关的文献较少, 尚无法对不同机型及扫描参数进行分层研究。DWI 的 ADC 值测量的临床应用价值仍有待于前瞻性的研究进一步证实。

参 考 文 献

- [1] 杨玲, 李连弟, 陈育德, 等. 中国乳腺癌发病死亡趋势的估计与预测. 中华肿瘤杂志, 2006, 28: 438-440.
- [2] The Cochrane methods group on systematic review of screening and diagnostic tests: recommended methods. [2006-04-07]. <http://www.Cochrane.Org/docs/sadtdocl.Htm>.
- [3] 王家良. 循证医学. 北京: 人民卫生出版社, 2001: 190.
- [4] Woodhams R, Matsunaga K, Kan S, et al. ADC mapping of benign and malignant breast tumors. Magn Reson Med Sci, 2005, 4: 35-42.
- [5] Rubesova E, Grell AS, De Maertelaer V, et al. Quantitative diffusion imaging in breast cancer: a clinical prospective study. J Magn Reson Imaging, 2006, 24: 319-324.
- [6] 郭勇, 蔡祖龙, 蔡幼铨, 等. 弥散加权成像鉴别乳腺良恶性病变的价值初探. 中华放射学杂志, 2001, 35: 132-135.
- [7] 罗建东, 刘圆圆, 张雪林, 等. 磁共振扩散加权成像在乳腺病变鉴别诊断中的应用价值. 癌症, 2007, 26: 168-171.
- [8] 李洁, 张晓鹏, 曹崑, 等. 乳腺 MR 动态增强扫描联合扩散加权成像的临床应用评价. 中国医学影像技术, 2005, 21: 1821-1825.
- [9] 赵斌, 蔡世峰, 高佩红, 等. MR 扩散加权成像鉴别乳腺良恶性病变的研究. 中华放射学杂志, 2005, 39: 497-500.
- [10] Guo Y, Cai YQ, Cai ZL, et al. Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging. J Magn Reson Imaging, 2002, 16: 172-178.
- [11] Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. BMJ, 2001, 323: 157-162.

(收稿日期: 2007-10-30)

(本文编辑: 任晓黎)