

抗肿瘤药品临床综合评价技术指南

(2021 年版)

国家药物和卫生技术综合评估中心

(国家卫生健康委卫生发展研究中心)

国家癌症中心

国家卫生健康委药具管理中心

二〇二一年十二月

目录

前 言.....	1
第一章 评价内容与维度.....	4
一、安全性评估.....	4
(一) 定义.....	4
(二) 指标选择.....	5
(三) 安全性测量.....	9
二、有效性评估.....	10
(一) 定义.....	10
(二) 指标选择.....	10
(三) 有效性测量.....	16
三、经济性评估.....	16
(一) 定义.....	16
(二) 临床结果指标选择与测量.....	17
(三) 成本指标选择与测量.....	20
(四) 经济学评估分析.....	24
四、创新性评估.....	28
五、适宜性评估.....	30
(一) 药品技术适宜性.....	31
(二) 药品使用适宜性.....	31
(三) 药品体系适宜性.....	31
(四) 药品监管适宜性.....	32
六、可及性评估.....	34
(一) 可获得性.....	34
(二) 可负担性.....	35
第二章 评价设计.....	37
一、评价技术类型.....	37
(一) 快速综合评价.....	37
(二) 完整综合评价.....	38
二、评价流程.....	39
三、抗肿瘤药品临床综合评价实施路径.....	40
(一) 评价设计.....	40
(二) 证据整合.....	43
(三) 评审决策.....	44
(四) 结果征求意见.....	44
(五) 应用转化.....	45
第三章 评价方法.....	46
一、基于一手资料的评价方法.....	46
(一) 真实世界数据.....	46
(二) 真实世界研究的基本步骤.....	49
(三) 偏倚的控制方法.....	49
(四) 真实世界药物经济学评价.....	53
(五) 数据信息安全.....	56

(六) 专家咨询法	56
二、基于二手资料的评价方法	58
(一) 确定研究问题	59
(二) 检索文献	60
(三) 建立纳入、排除标准	61
(四) 证据评价与证据整合	62
(五) 质量控制	62
第四章 基于多准则决策分析的药品临床综合评审流程	63
一、建立专家组	64
二、抗肿瘤药物信息介绍	64
三、维度和指标审阅	64
四、维度与指标赋权	65
五、指标赋分	65
六、分值计算	66
七、形成推荐意见	66
八、综合评判	67
参考文献	68
中英文对照表	69
《抗肿瘤药品临床综合评价技术指南》编委会	71
附件 标准报告格式	72

前 言

恶性肿瘤(癌症)已成为严重威胁我国人群健康的主要致死疾病之一。根据 2020 年 Globocan 统计,我国癌症新发病例约为 456.88 万人;年龄标化(全球)发病率男性为 225.4/10 万,女性为 188.2/10 万。2020 年我国癌症死亡病例约为 300.29 万人,其中男性为 182.0 万人,女性为 118.3 万人^[1]。

恶性肿瘤治疗手段发展较快。随着肿瘤生物学研究的快速进展和对恶性肿瘤疾病发病机理认识的不断深入,抗肿瘤药品的研发创新不断取得突破性进展,特别是新作用机制、作用靶点的抗肿瘤药品不断涌现上市或进入临床试验。加快建立健全统一、科学、实用的药品临床综合评价标准规范、实施路径和工作协调机制成为当前首要工作任务之一。

依据国家药品临床综合评价管理政策和制度规范,为指导和规范卫生健康系统开展抗肿瘤药品临床综合评价,国家卫生健康委药物政策与基本药物制度司(以下简称“国家卫生健康委药政司”)委托国家药物和卫生技术综合评估中心(以下简称“评估中心”)(挂靠国家卫生健康委卫生发展研究中心)联合国家癌症中心、国家卫生健康委药具管理中心(以下简称“药具中心”)等,成立指南起草小组,听取全国医药行业学(协)会意见,组织全国医学、药学、临床研究、循证医学、卫生经济学、卫生政策等领域专家共同制定本技术指南。

本技术指南依据国家《药品临床综合评价管理指南（2021年版 试行）》（以下简称《管理指南》）的组织管理及实施流程要求，根据我国抗肿瘤药品临床应用实践与药物供应保障政策现况，重点聚焦我国医疗机构抗肿瘤药品临床使用和技术评价存在的实际问题，参考借鉴国际抗肿瘤药品评估有益做法和经验，旨在引导和推动相关主体规范开展抗肿瘤药品临床综合评价，持续推动抗肿瘤药品临床综合评价工作标准化、规范化、科学化、同质化，更好地服务国家抗肿瘤药品政策决策需求，助力提高药事服务质量，控制不合理药品费用支出，保障临床抗肿瘤药品的供应与规范使用。

本指南基于当前中国实施药品临床综合评价的认知、技术能力和资源动员能力编制，主要供涉及药品临床综合评价研究和实施的相关主体依规定知情使用。随着理论研究的深入、科技发展进步和实践证据的积累改进，本指南将不断进行更新与完善。

适用范围 本指南用于国家及省级抗肿瘤药品临床综合评价的技术指导，同时为医疗卫生机构、科研院所、大专院校、行业学（协）会等主体开展抗肿瘤药品临床综合评价相关研究提供技术规范和流程指引，并为开展抗肿瘤中药的临床综合评价提供参考借鉴。

解释权 评估中心对本指南的技术应用部分具有解释权。

修订期限 本指南采用开放式动态修订机制，长期接受

来自社会各界的建议，并定期组织专家对有建设性意见进行研究和讨论。如对本指南的任何意见或建议，请联系 NCHTA@nhei.cn。

指南获取 国家卫生健康委卫生发展研究中心官网（官方网址：<http://www.nhei.cn>）和国家癌症中心官网（官方网址：<http://www.cicams.ac.cn/Html/News/Main/249.html>）

第一章 评价内容与维度

抗肿瘤药品临床综合评价是评价主体应用多种评价方法和工具开展的多维度、多层次证据的综合评判。评价主要聚焦抗肿瘤药品临床使用实践中的重大技术问题和政策问题，围绕技术评价与政策评价两条主线，从安全性、有效性、经济性、创新性、适宜性、可及性 6 个维度开展科学规范的定性定量相结合的证据整合分析与综合研判，提出国家、区域或省级层面临床重大疾病防治基本药物供应与规范使用的政策建议。

一、安全性评估

（一）定义

安全性评估指对抗肿瘤药品质量及上市前后出现的或可能出现的用药风险进行科学性评估。用药风险包括药品不良反应以及与用药有关的不良事件发生风险等。安全性是抗肿瘤药品临床综合评价的核心维度，是判断药品临床价值的基础。除特定情况，建议所有针对抗肿瘤药品的临床综合评价均应开展药品临床使用的安全性评估。

同类药品安全性比较。是指与同适应症对照药比较的安全性差异，即不良反应发生率和/或严重程度的差异，重点采用真实世界研究、文献评价等方法，关注严重不良事件发生率的比较。针对真实世界研究建议参考相关指南规范，综合利用医疗机构、疾病登记系统等数据开展研究，全面考察真实世界中药品近、远期安全性差异、不同患者群体的安全性差异等。

(二) 指标选择

评估指标可选药品不良事件发生率，主要依据常见不良反应事件评价标准（Common Terminology Criteria for Adverse Events, CTCAE）（见表1），CTCAE将药品不良事件按不同系统、不同级别分为五级，安全性评估需考虑总体不良事件发生率和三级及以上不良事件发生率两类指标。

表1 常见不良事件评价标准（CTCAE 5.0版）节选

不良事件	分级				
	1	2	3	4	5
血液和淋巴系统					
贫血	血红蛋白<正常值下限-10.0g/dL; <正常值下限-6.2mmol/L; <正常值下限-100g/L	血红蛋白<10.0-8.0g/dL; <6.2-4.9mmol/L; <100-80g/L	血红蛋白<8.0g/dL; <4.9mmo/L ; <80g/L; 需要输血治疗	危及生命; 需要紧急治疗	死亡
骨髓细胞过少	轻微细胞过少或与该年龄段的正常细胞总数相比减少≤25%	中度细胞过少或与该年龄段的正常细胞总数相比减少>25%且≤50%	重度细胞过少或与该年龄段的正常细胞总数相比减少>50%且≤75%	再生障碍持续2周以上	死亡
发热性中性粒细胞减少	-	-	ANC<1000/mm ³ 伴有单次体温>38.3°C(101°F)	危及生命; 需要紧急治疗	死亡

不良事件	分级				
	1	2	3	4	5
			或 体 温 ≥38°C(100.4°F) 持续1小时以上		
胃肠道					
便秘	偶然或间断性出现; 偶尔需使用通便剂、轻泻药、调整饮食习惯或灌肠	持续症状, 需要使用规律的轻泻药或灌肠; 借助于工具的生活活动受限	需手工疏通的顽固性便秘; 自理性日常生活受限	危及生命; 需紧急治疗	死亡
腹泻	与基线相比, 大便次数增加, 每天<4次; 造轻度增加	与基线相比, 大便次数增加, 每天4~6次; 造中度增加; 借助于工具的生活活动受限	与基线相比, 大便次数增加, 每天≥7次; 需住院治疗; 造重度增加; 自理性日常生活活动受限	危及生命; 需紧急治疗	死亡
胃炎	无症状; 仅为临床或诊断所见; 无需治疗	有症状; 肠道功能改变; 需治疗	进食或胃功能重度改变; 需肠外营养或住院治疗	危及生命; 需紧急手术	死亡
口腔黏膜炎	无症状或轻度; 不需要治疗	中度疼痛或溃疡; 影响进食; 需要调整饮食	重度疼痛; 影响经口进食	危及生命; 需紧急治疗	死亡
恶心	食欲降低, 不伴进食习惯改变	经口摄食减少, 不伴明显的脱水或营养不良	经口摄入量不足; 需鼻饲、全肠外营养或者住院	-	-
呕吐	不需要进行干预	门诊静脉补液; 需进行医学干预	需要鼻饲、全肠外营养或住院治疗	危及生命	死亡
全身性					
疲劳	疲劳, 休息后可缓解	疲劳, 休息后不能缓解; 影响日常家务活动	疲劳, 休息后不能缓解; 影响自理性日常生活活动	-	-
发热	38.0~39.0°C(100.4-102.2°F)	>39.0~40.0°C(102.3-104.0°F)	>40.0°C(>104.0°F)≤24小时	>40.0°C(>104.0°F) 超过24小时	死亡
疼痛	轻度疼痛	中度疼痛; 影响日常家务活动	重度疼痛; 影响自理性日常生活活动	-	-
免疫系统					
变态反应	不需要系统干预	需要口服给药治疗	支气管痉挛; 需住院治疗; 临床后遗症; 需静脉干预治疗	危及生命; 需紧急治疗	死亡
感染和传染					
上呼吸道感染	-	中度; 需要口服药物 (抗生素, 抗真菌或抗病毒治疗)	需静脉注射抗生素, 抗真菌或抗病毒药物	危及生命; 需紧急治疗	死亡

不良事件	分级				
	1	2	3	4	5
			治疗; 需要进行侵入性的治疗		
医学检查					
丙氨酸氨基转移酶增高	>正常值上限的3倍(基线值正常); 基线值的1.5-3.0倍(如基线值不正常)	正常值上限的3-5倍(基线值正常), 大于基线值的3.0-5.0倍(如基线值不正常)	5.0-20.0倍(基线值正常); 大于基线值的5-20倍(如基线值不正常)	大于正常值上限20倍(基线值正常), 大于基线值20倍(如基线值不正常)	-
碱性磷酸酶增高	大于正常值上限的2.5倍(基线值正常); 基线值的2-2.5倍(基线值不正常)	大于正常值上限的2.5-5.0倍(基线值正常); 基线值的2.5-5.0倍(基线值不正常)	大于正常值上限的5.0-20倍(基线值正常); 基线值的5.0-20倍(基线值不正常)	大于正常值上限20倍(基线值正常), 大于基线值20倍(如基线值不正常)	-
血胆红素增高	>1.5倍正常值上限(基线值正常); 大于1-1.5倍基线值(基线值不正常)	大于1.5-3.0倍正常值上限(基线值正常); 大于1.5-3.0倍基线值(基线值不正常)	大于3.0-10倍正常值上限(基线值正常); 大于3.0-10倍基线值(基线值不正常)	大于10倍正常值上限(基线值正常); 大于10倍基线值(基线值不正常)	-
INR增高	>1.2-1.5; >1-1.5倍基线水平(抗凝时); 只需监测	>1.5-2.5; >1.5-2.5倍基线水平(抗凝时); 提示剂量调整	>2.5; >2.5倍基线水平(抗凝时); 出现出血	-	-
心电图QTc间期延长	QTc 450-480ms	QTc 481-500ms	QTc≥501ms; 比基线期>60ms	尖端扭转型室速; 阵发性室性心动过速; 严重心律不齐征/症状	-
骨骼肌和结缔组织					
关节痛	轻度疼痛	中度疼痛; 影响工具性日常生活活动	重度疼痛; 影响自理性日常生活活动	-	-
骨痛	轻度疼痛	中度疼痛; 影响工具性日常生活活动	剧痛; 影响自理性日常生活活动	-	-
肌痛	轻度疼痛	中度疼痛; 影响工具性日常生活活动	剧痛; 影响自理性日常生活活动	-	-
神经系统					
头痛	轻度疼痛	中度头痛; 影响工具性日常生活活动	重度头痛; 影响自理性日常生活活动	-	-
外周运动神经障碍	无症状, 仅为临床或诊断所见	中度症状; 影响工具性日常生活活动	重度症状; 影响自理性日常生活活动	危及生命; 需要紧急治疗	死亡
外周感觉神经障碍	无症状	中度症状; 影响工具性日常生活活动	重度症状; 影响自理性日常生活活动	危及生命; 需要紧急治疗	-
肾脏和泌尿					

不良事件	分级				
	1	2	3	4	5
急性肾损伤	-	-	需要住院治疗	危及生命; 需要透析治疗	死亡
慢性肾脏疾病	肾小球滤过率或肌酐清除率小于60ml/min/1.73m ² 或蛋白尿2+; 尿蛋白/肌酐大于0.5	肾小球滤过率或肌酐清除率59-30ml/min/1.73m ²	肾小球滤过率或肌酐清除率29-15ml/min/1.73m ²	肾小球滤过率或肌酐清除率小于15ml/min/1.73m ² ; 需要透析或肾移植	死亡
血尿	无症状, 仅为临床或诊断所见; 无需治疗	有症状; 需要导尿管或膀胱冲洗; 影响工具性日常生活活动	大量血尿; 需要输血, 静脉给药或住院治疗; 需要择期侵入性治疗; 个人活动能力受限	危及生命; 需要紧急侵入性治疗	死亡
蛋白尿	蛋白尿1+, 24小时尿蛋白小于1.0g	成人: 蛋白尿2+和3+, 24小时尿蛋白1.0-3.5g 儿童: 尿液中蛋白/肌酐比值0.5-1.9	成人: 24小时尿蛋白≥3.5g, 4+蛋白尿 儿童: 尿蛋白/肌酐比值大于1.9	-	-
呼吸、胸部和纵隔					
咳嗽	轻度症状; 需要非处方治疗	中度症状; 需要药物治疗; 影响工具性日常生活活动	重度症状; 影响自理性日常生活活动	-	-
呼吸困难	中度活动时呼吸短促	少量活动时呼吸短促; 影响工具性日常生活活动	休息时呼吸短促; 影响自理性日常生活活动	危及生命; 需要紧急干预	死亡
呃逆	轻度症状; 无需治疗	中度症状; 需要治疗; 影响工具性日常生活活动	重度症状; 影响睡眠; 影响自理性日常生活活动	-	-
皮肤和皮下组织					
脱发	个体脱发小于50%, 远距离观察, 无明显区别, 但近距离观察可见。需要改变发型来掩饰, 但不需要假发或假发来掩饰	个体脱发大于等于50%, 症状明显; 如果患者想要完全掩盖头发丢失, 需要假发或假发片, 伴心理影响	-	-	-
皮肤干燥	累及小于10%体表面积, 无红斑或瘙痒	累及10%~30%体表面积, 伴有红斑或瘙痒; 影响工具性日常生活活动	累及超过30%体表面积, 伴有瘙痒; 影响自理性日常生活活动	-	-
瘙痒症	轻度或局限性的; 需要局部的治疗	广泛分布且间歇性发作; 搔抓引起皮肤改变(肿胀, 丘疹, 脱	广泛分布且持续性发作; 影响自理性日常生活活动或睡眠;	-	-

不良事件	分级				
	1	2	3	4	5
		皮，苔藓样，渗出/结痂；需要口服药治疗；影响工具性日常生活活动	需要全身性糖皮质激素或免疫抑制剂治疗		
皮肤色素沉着过多	色素沉着小于10%的体表面积；没有心理影响	色素沉着大于10%的体表面积；伴有心理影响	-	-	-

注:等级是指不良事件的严重程度。CTCAE 根据下面的一般准则对各个不良事件的严重程度(1 级至 5 级)作了特定的临床描述:

1 级: 轻度; 无症状或轻微症状; 仅为临床或诊断所见; 无需治疗。

2 级: 中度; 需要较小、局部或非侵入性治疗; 与年龄相当的工具性日常生活活动受限。

3 级: 严重或者具有重要医学意义但不会立即危及生命; 导致住院或者延长住院时间; 致残; 自理性日常生活活动受限。

4 级: 危及生命; 需要紧急治疗。

5 级: 与 AE 相关的死亡。

日常生活活动(Activities of Daily Living, ADL):

工具性日常生活活动指做饭、购买衣物、使用电话、理财等。

自理性日常生活活动指洗澡、穿脱衣、吃饭、盥洗、服药等, 并未卧床不起。

(三) 安全性测量

推荐首选国内药监部门提供的药品质量及不良反应监测结果, 其次应获取临床药品使用安全数据等, 还可综合参考文献证据的系统评价/Meta 分析。

建议充分利用国家药监局、药政部门已有的药品信息系统数据, 包括药品上市许可持有人、生产企业、经营企业及临床用药机构提供的信息及数据。

基于一般可获得资料信息及数据来源(见表 2), 结合具体评价药品及主要评估问题选择适宜的安全性指标信息来源。

表 2 安全性评估信息来源

安全性信息	具体内容	来源
上市前安全性信息	不良反应、耐受情况、禁忌症、注意事项、特殊人群用药、药物相互作用、药物过量等	药品说明书、临床试验研究结果，药品审批申报材料
上市后不良事件及不良反应发生率	上市后临床安全性和报告中任何不良反应发生率	国家药品不良反应监测中心数据库；文献系统评价/Meta 分析，药监局公告等信息，合理用药国际网络，临床安全用药监测网，医院病案数据、企业用药监测数据
同类药品安全性比较	比较同类或同一适应证不同类别药品的安全性	文献数据库，专家/患者咨询

二、有效性评估

（一）定义

有效性评估是指抗肿瘤药品使用后对患者疾病状况的治疗程度进行科学评估。

（二）指标选择

抗肿瘤药品有效性评估的指标主要以患者用药后的实际疗效为依据，设立主要临床结局、次要临床结局和患者报告结局指标。优先选取能够反映患者长期获益及用药后整个生命进程疾病转归等相关指标为主要临床结局，其他可准确测量的临床指标等作为次要临床结局。同时，肿瘤患者可能因疾病本身以及治疗相关的不良反应而严重影响其生活质量，本指南纳入患者报告结局指标，反应患者声音。

1. 主要临床结局指标

主要临床结局指标主要包括总生存期与无进展生存期。

(1) 总生存期 (Overall survival, OS)：从界定的开始时间 (如随机分组) 至因各种原因导致病人死亡的时间。

(2) 无进展生存期 (Progression-free survival, PFS)：从界定的开始时间 (如随机分组) 至出现肿瘤客观进展或死亡的时间。

若同时有总生存期和无进展生存期，推荐使用总生存期作为主要临床结局指标。若使用无进展生存期作为主要临床结局，需详细阐述原因。

2. 次要临床结局指标

次要临床结局指标主要包括至疾病进展时间、无病生存期、无远处转移生存期和客观缓解率。

(1) 至疾病进展时间 (Time To Progression, TTP)：从界定的开始时间 (如随机分组) 至出现肿瘤客观进展的时间，不包括死亡。

(2) 无病生存期 (Disease-free Survival, DFS)：从界定的开始时间 (如随机分组) 至出现肿瘤复发或由任何原因引起死亡的时间。

(3) 无远处转移生存期 (Distance Disease-free Survival, DDFS)：经过治疗后除原发病灶以外，未发现全身其他地方有转移病灶的时间。

(4) 客观缓解率 (Objective Response Rate, ORR): 肿瘤体积缩小达到预先规定值并能维持最低时限要求的患者比例。对肿瘤负荷进行评估是判断肿瘤治疗效果的重要手段之一, 瘤体皱缩 (目标疗效) 和病情恶化在临床试验中都是有意义的判断终点, 评价标准可依据实体瘤疗效评价标准指南 (Response Evaluation Criteria in Solid Tumors, RECIST) (见表 3)。

表 3 实体瘤疗效评价标准指南 (1.1 版)

基线病灶分类	定义
可测量病灶	至少有一条可以精确测量的径线的病灶。 根据断层扫描 (CT) 或核磁共振成像 (MRI) 评价, 病灶最长直径至少为 2 个层厚 $\geq 10\text{mm}$ (层厚 5-8mm)。 根据胸部 X 线评价, 病灶最长直径至少 20mm。 根据测径器评价, 最长直径 $\geq 10\text{mm}$ 的浅表性病灶。 根据 CT 评价, 恶性肿瘤淋巴结短轴 $\geq 15\text{mm}$ 。 注: 恶性肿瘤淋巴结用最短轴作为直径, 其它可测量病灶用最长轴。
不可测量病灶	不可测量病灶包括小病灶 (包括短轴在 10-14.9mm 的淋巴结) 和真正无法测量的病灶, 如胸膜或心包积液、腹水、炎性乳腺疾病、软脑膜病、累及皮肤或肺的淋巴管炎, 测径器不能准确测量的临床病灶, 体检发现的腹部肿块, 重现影像技术无法测量的。 骨病: 骨病为不可测量的疾病, 除软组织成份可采用 CT 或 MRI 评价外, 且符合基线时可评价的定义。 既往局部治疗: 既往放疗病灶 (或其它局部治疗的病灶) 为不可测量病灶, 除非治疗完成后进展。
正常部位	囊性病灶: 单纯囊肿不应视为恶性病灶, 也不应记录为目标病灶或非目标病灶。认为是囊性转移的囊性病灶是可测量病灶, 如果符合上述特定定义。如果还出现了非囊性病灶, 那么这些病灶首选为目标病灶。 正常结节: 短轴 $< 10\text{mm}$ 的结节被视为正常, 不应记录或安装可测量或不可测量病灶分类。
记录治疗评价	在基线时必须评价所有部位疾病。基线评价应尽量在接近试验开始前进行。对于充分的基线评价, 治疗前 28 天内必须进行所有要求的扫描, 所有疾病必须正确记录。如果基线评价不充分, 以后的状况通常为不确定。
目标病灶	所有累及器官达最多 2 个病灶/每个器官, 共 5 个病灶, 所有可测量病灶应视为基线目标病灶。根据大小 (最长病灶) 和适合性选择目标病灶准确重复测量。记录每个病灶的最长直径, 除外病理学淋巴结应记录短轴。基线时所有目标病灶直径 (非结节病灶的最长径, 结节病灶的最短轴) 的总和是试验中进行的评价比较的基础。 若两个病灶融合, 就测量融合的肿块。如果目标病灶分裂, 则使用各部分的总和。

基线病灶分类	定义
	<p>·应继续记录变小的目标病灶的测量。如果目标病灶变的太小而不能测量，如认为病灶已消失则记录为 0mm；反之应记录为默认值 5mm。 注：结节性标准缩小至 < 10mm（正常），仍应记录实际测量结果。</p>
非目标病灶	<p>所有不可测量的疾病均为非目标病灶。所有未鉴别为目标病灶的可测量病灶也纳入非病灶疾病。不需要进行测量，但是评价以无、不确定、有/未增大、增大表示。1 个器官的多发性非目标病灶在病例报告表上记录为一项（如：‘多发性骨盆淋巴结增大’或‘多发性肝转移’）。</p>
每次评价时的客观缓解状态	<p>疾病部位评价必须采用与基线相同的方法，包括一致进行增强和及时的扫描。如需变化，必须与放射学医师讨论该病例以明确是否可能用替代法。如不能，以后的客观状况为不明确。</p>
目标病灶	<p>完全缓解（Complete Response, CR）：除结节性疾病外，所有目标病灶完全消失。所有目标结节须缩小至正常大小（短轴 < 10mm）。所有目标病灶均须评价。 部分缓解（Partial Response, PR）：所有可测量目标病灶的直径总和低于基线 ≥ 30%。目标结节总和使用短径，而所有其它目标病灶的总和使用最长直径。所有目标病灶均须评价。 稳定（Stable Disease, SD）：不符合 CR、PR 或进展。所有目标病灶均须评价。仅在总增大相对于谷值 < 20% 的罕见病例，PR 后可稳定，但足够不再维持以前记录的缩小 30%。 疾病进展（Progression Disease, PD）：可测量目标病灶的直径总和增大 20% 超过观察到的最小总和（超过基线，如治疗期间未观察到总和降低），最小绝对值升高 5mm。 不确定：未记录进展，且 1 个或以上可测量的目标病灶未评价或所用评价方法与基线不一致；或 1 个或以上目标病灶不能准确测量（如：看不清除，除非由于太小而不能测量）；或 1 个或以上目标病灶被切除或辐射，且未复发或增大。</p>
非目标病灶	<p>CR：所有非目标病灶消失或肿瘤标志物水平正常。所有淋巴结大小必须‘正常’（短轴 < 10mm）。 非 CR/非 PD：任何非目标病灶持续存在和/肿瘤标志物水平高于正常上限。 PD：已有病灶明确进展。通常，总体肿瘤负荷须增大到足以停止治疗。目标病灶 SD 或 PR 时，罕见由于非目标病灶明确增大的进展。 不明确：未测量进展，1 个或以上非目标病灶部位未评价或评价方法与基线所用方法不一致。</p>
新病灶	<p>出现任何新发明明确的恶性肿瘤病灶都表明 PD。如果新病灶不明确，例如由于体积较小，进一步评价会明确病因。如果重复评价明确病灶，那么应在首次评价日期记录进展。在以前未扫描区发现的病灶被认为是新病灶。</p>
补充研究	<p>如果明确 CR 取决于体积减小但未完全消失的残留病灶，建议活检或细针抽吸残留病灶进行研究。如未发现疾病，主观状况记录为 CR。 如果明确进展取决于可能由于坏死增大的病灶，那么病灶应活检或细针抽吸以明确状态。</p>
主观性进展	<p>无疾病进展的客观证据，需要终止治疗的患者，在肿瘤评价癌症相关性乏力 (Cancer-Related Fatigue Symptoms, CRFS) 上不应报告为 PD。这一情况应在治疗结束 CRFS 上标明为由于健康状况总体恶化停止治疗。即使在停止治疗后也应尽量记录客观进展。</p>

RECIST 标准在近 10 余年应用过程中几经更新，已成为国内外肿瘤界采用的通用标准。然而随着介入治疗、靶向治疗以及免疫治疗药物在临床的实际应用，新评价标准将肿瘤大小和密度相结合，以满足传统实体瘤放射、化学治疗的效果评价标准无法准确而全面地评价介入、靶向及免疫治疗药物的抗肿瘤效果。新评价标准如基于 RECIST 衍生的其他相关标准（mRECIST、iRECIST）、胃肠道间质瘤的 CT 疗效评价改良标准（Choi 标准）和免疫相关评价标准（Immune-related response criteria, irRC）等。上述列举标准各有利弊，在实际抗肿瘤药品临床综合评价中，应结合被评价药物的实际应用特点，酌情选择适宜的评价标准。

对于纳入研究的患者，都应进行相应疗效评价，其分类为（1）完全缓解；（2）部分缓解；（3）疾病稳定；（4）疾病进展；（5）因肿瘤致早期死亡；（6）因治疗毒性致早期死亡；（7）其他原因致早期死亡；（8）无法分类（不能评价或资料不完整），具体结果判定（见表 4）。

表 4 总疗效评价

目标病灶	非目标病灶	新病灶	总疗效
CR	CR	No	CR
CR	SD	No	PR
PR	Non-PD	No	PR
SD	Non-PD	No	SD
PD	Any	Yes/No	PD
Any	PD	Yes/No	PD
Any	Any	Yes	PD

血液肿瘤主要包括白血病、淋巴瘤和骨髓瘤三大类。中国淋巴瘤治疗指南（2021年版）指出目前主要采用Lugano2014评价标准评价淋巴瘤的治疗效果，治疗效果分为基于CT和（或）MRI评价的影像学缓解和基于PET-CT评价的代谢缓解，PET-CT评价代谢缓解的依据是PET5分法（Deauville标准）。白血病和骨髓瘤可分多种临床类型，不同类型的疗效评价标准也不完全相同，具体评价时可参考相关病种指南或诊疗规范，如美国国立综合癌症网络（National Comprehensive Cancer Network, NCCN）指南，肿瘤诊疗规范等。

3. 患者报告结局指标

患者报告结局（patient-reported outcome, PRO）是直接来自于患者对自身健康相关生活质量的报告，主要包括患者对自身健康状况、功能状态以及治疗感受的报告。肿瘤患者报告结局关注抗肿瘤药物治疗对患者健康相关生活质量的影响，包括疾病相关的症状、症状相关的不良事件、生理功能、心理功能及社会角色。患者报告结局测量方式包括患者自填量表或者问卷、面对面定性访谈以及电话访谈等。推荐使用基于中国人群的患者报告结局量表进行数据收集，包括通用量表和肿瘤疾病别特异性量表。量表具体介绍见经济性评估章节。

(三) 有效性测量

有效性指标的测量应该选择可获得的最佳证据，即从临床疗效或实际效果指标中选优。对于新药，当临床试验的疗效数据可获得并适用时，优先选择临床疗效数据；对于已上市多年的药品，建议考虑使用真实世界研究中的实际效果数据。

对于实效性临床研究(Pragmatic randomized clinical trial, PRCT)和RCT数据的选择，推荐优先考虑基于中国人群开展的大样本、多中心数据或含有中国人群的国际多中心数据。如选择含有中国人群的国际多中心RCT数据，应尽可能地对亚组人群的特征进行描述和分析；当无法获得含有中国人群的数据时，可以使用基于其他国家或地区人群的数据，但需要清晰地解释使用这些数据的合理性，强调人群之间潜在的差异，并需要对关键参数进行敏感性分析。

推荐优先选择待评价药品与对照药品头对头直接比较的RCT和PRCT数据。当缺少直接比较时，应尽量选择间接比较或网络荟萃分析的数据。

三、经济性评估

(一) 定义

经济性评估主要用于评价资源消耗与临床疗效的性价

比。资源消耗可通过费用指标数据与成本指标数据进行衡量；健康产出可以通过效果指标数据和效用指标数据进行衡量；衡量其临床使用的经济价值及预算影响^[2]。经济性评估方法建议优先考虑成本-效用分析（Cost-Utility Analysis, CUA）和成本-效果分析（Cost-Effectiveness Analysis, CEA）。可通过模型分析法评估药品间的增量资源消耗和增量健康产出比值，进而形成优选方案。经济性评估实施主要参考现有药物经济学评价指南，主要包括中国药物经济学评价指南、中国药品综合评价指南参考大纲以及国际药物经济学和结果研究协会等发表的系列指南。

（二）临床结果指标选择与测量

1. 效果/疗效

效果/疗效指标指抗肿瘤药品在真实临床应用环境中的实际效果/疗效，包括主要临床结局指标，例如总生存期（OS）、无进展生存期（PFS）；次要临床结局指标，例如至疾病进展时间（TTP）、客观缓解率（ORR）等。效果指标还可为住院天数、住院死亡率、重复入院率等（可通过HIS信息系统获得）。指标的采集可通过一手资料法从医疗机构临床实践数据中获取（如HIS信息系统与患者随访数据），或来源于二手文献资料。

2. 效用

在无法开展药品等效评价的情况下，推荐使用质量调整生命年（Quality-Adjusted Life Years, QALYs）作为效用指标开展评价。QALYs 的计算为延长的生命年数乘以这段时间内的健康效用值（权重值），既考虑了治疗方案对患者生存时间的影响，也考虑了对患者生命质量的影响。

效用值的选择首先推荐基于中国人群建立的健康效用积分体系进行效用值的计算。当没有基于中国人群的健康效用积分体系时，可采用其他普遍认可并广泛应用的健康效用积分体系，但需阐明其适用性，并进行敏感性分析。若无可获得效用数据，推荐通过调研测量相关疾病效用值。效用值可用通用量表或疾病专用量表间接测量获得，常用通用量表有欧洲五维健康量表（EuroQoL Five Dimensions Questionnaires, EQ-5D）等；肿瘤常用疾病专用量表主要由系列生活质量量表（Quality of Life Core Questionnaire, QLQ）组成。QLQ 为系列量表，由针对所有癌症患者的核心量表（共性模块）QLQ-C30 和针对不同癌症的特异性条目（特异模块）构成的量表群组成，拥有肺癌（QLQ-CL31）、乳腺癌（QLQ-B2R3）、头颈癌（QLQ-HN35）、直肠癌（QLQ-C38）等多个特异性模块。

当无可用间接测量量表时，可采用直接测量法测量效用值。直接测量量表包括视觉模拟标尺法、标准博弈法、时间权衡法等。推荐直接通过患者本人测量其健康相关生命质量。

当无法通过测量得到效用值时，可从文献中获取相似问题、相似人群及状态的效用值，但需进行敏感性分析。

3. 效益

效益是指用货币单位量化的健康产出。疾病治疗方案的效益包括直接效益、间接效益和无形效益三个部分。

直接效益指实行某项干预措施后所节省的卫生资源，可以用货币计量直接效益。

间接效益指实行某项干预措施后所增加的患者健康时间或劳动生产力恢复带来的收益。推荐使用人力资本法计算间接效益。

无形效益指实行某项干预措施后减轻或者避免患者身体和精神上的痛苦，以及康复后带来的舒适和愉快等。推荐采用意愿支付法计算无形效益。

4. 健康产出贴现

当待评药品治疗时间超过一年时，应该对发生在未来的健康产出进行贴现。本指南建议贴现率可采用研究当年的 1 年期国债收益率，若采用其他贴现率，应给出合理解释。

5. 临床结果外推

效果外推主要涉及的问题有两个方面，一是在严格的临床试验条件下得到的疗效是否能够反映在更广泛的“真实世界”中的治疗效果。由于临床试验的样本纳入或排除标准十分严格且细致，其结果的外推性会受到一定程度的影响；二是

真实世界下得到的效果是否能直接运用到不同的临床诊疗环境之中。需分析评价纳入的目标人群的特征（人口特征、疾病病情特征、所属地区医疗保障体系特征、文化特点等）与外推人群特征的异同，判断外推的合理性和可行性。

（三）成本指标选择与测量

1. 成本定义及分类

药物治疗成本是指因接受抗肿瘤药品治疗而消耗的医疗资源和相关非医疗资源。因此，药品治疗成本不仅包括药品费用本身，还包括相关的诊查、随访、照护等临床治疗产生的直接或间接费用。

药物治疗成本主要包括直接成本、间接成本和隐性成本，对于间接成本和隐性成本，研究者可以灵活处理，但要避免重复计算；直接成本包括直接医疗成本和直接非医疗成本，直接医疗成本通常包括门诊费用、住院费用、自购药费用等；直接非医疗成本通常包括交通费、营养费、陪护费等；间接成本指由于疾病、伤残或死亡造成的患儿和其家庭的劳动时间及生产率损失，包括休学、停工、早亡等造成的患儿及家人的经济收入损失等。当隐形成本显著较大时，需要对其进行专门的评估。

2. 成本测量的角度

成本范围的界定需要与所确定的研究角度和研究时间保持一致。常用的研究角度有：全社会角度、卫生体系角度、医疗保障支付方角度、医疗机构角度和患者角度等。

全社会角度下应纳入所有直接医疗成本、直接非医疗成本和间接成本，所有应用于公共决策的经济学评价都应该提供全社会视角的研究结果，然而实践中全社会视角的研究结果较难获得；卫生体系角度下应纳入卫生系统内的所有直接医疗成本，目前我国抗肿瘤药品的综合评价工作由国家卫生健康部门牵头，一般情况下推荐采取卫生体系角度开展成本测量；医疗保障支付方角度下应纳入医保支付范围内的所有直接医疗成本；医疗机构角度下，应纳入在本医疗机构承担的直接医疗成本和非医疗成本（如果适用）；患者角度下，应纳入患者相关的所有直接医疗成本、直接非医疗成本和间接成本。

推荐采用卫生体系角度对抗肿瘤药品进行成本分析。抗肿瘤药品一般价格昂贵，是构成直接医疗成本的主要部分，应准确收集药品价格（考虑患者援助计划）、用法、用量、疗程等数据对药品费用进行测算。此外，抗肿瘤药品一般不良事件较多且涉及范围广，应关注不良事件产生的治疗费用，将其干预和监测成本纳入直接医疗成本。报告结果时建议在不同药品价格水平和不同情境下做治疗成本的情境分析，供

决策者参考。可采用国际上比较通用的生产要素法以及作业成本法 (Activity-Based Costing, ABC)进行成本分析。

3. 成本测量及测量范围

成本测量时应首先列出与实施干预措施相关的资源项目，并明确资源项目的计量单位，再根据该计量单位测算消耗的资源数量。计量单位主要包括两类，一是卫生资源消耗的自然单位，二是根据国家相关部门制定的项目标准。成本的计量单位可以是比较宏观的，如一年就诊、一次住院、一次门诊等单位；也可以是比较微观的，如一片药品、一次注射、一次护理等。在数据可得的情况下，应尽可能使用微观的计量单位。

成本测量的范围需要与所确定的研究时限一致，应纳入研究时限内与实施干预措施相关的所有当前的和未来的成本。成本测量的范围还应坚持与临床效果对等原则，要将产生“该类临床效果”对应的所有诊疗活动所产生的成本均纳入成本测量范围。

在基于临床试验的抗肿瘤药品临床综合评价中，应当识别并排除为了进行临床试验而发生、但在实际临床治疗中不会发生的“试验成本”。如某些成本项目难以确定在实际临床治疗中是否会发生，可以参考同类非基于临床试验的抗肿瘤药品临床综合评价中的成本构成进行敏感性分析。

如果待评药品发生了药品不良反应，则需计入待评药品

成本。与不良反应相关的成本主要有两类，一是为避免或监测不良反应发生而产生的成本，二是不良反应发生后进行医疗干预而产生的成本。尤其需要关注中度和重度不良反应对成本产生的影响。

对于因疾病治疗所付出的间接成本，也称劳动力成本，建议采用人力资本法进行计算，即假定所有损失的时间用于生产，用劳动力市场平均工资水平去估算因疾病或过早死亡带来的劳动力损失。

4. 药品价格

药品价格是成本测算的基础数据。药品单价建议优先使用官方或权威机构发布的最新价格信息（集中采购的中标价等），其次可选医疗卫生机构或正规经营药店的销售价格。如果某一医疗资源项目在市场上存在多个价格，在市场份额分布已知前提下，可采用市场份额加权的平均价格；在市场份额不可获得时，可采用所有已知价格的中位数或平均数。

同一通用名相同剂型不同规格药品的价格可以采用限定日剂量（Defined Daily Dose, DDD）进行校正，再进行加权或中位价格的计算；不同剂型的药品如果治疗目标相当，也可以通过 DDD 方式进行价格转换，然后再计算加权平均价格或中位价格。

如果药品尚未在国内上市，建议采用生产厂商建议价格，

同时获取多个国家及地区的国际价格作为参考^[3]，确定适宜的价格或价格区间。研究者如未使用上述推荐的价格体系，应该明确注明并解释所使用的其它价格体系的合理性。

5. 成本贴现

当待评药品治疗时间超过一年时，应该对发生在未来的成本进行贴现。建议对成本与健康产出采用相同的贴现率。

6. 成本外推

不同国家和地区之间的医疗资源使用标准和单位成本之间存在较大差异，成本外推时需分析本次评价目标人群所在地区的经济水平、各类相关资源的价格水平与外推地区的异同，判断外推的合理性和可行性。

（四）经济学评估分析

根据评估决策需求，对待评主题进行经济学评估。评估方法主要有成本-效果分析(CEA)、成本-效用分析(CUA)、成本-效益分析(CBA)、最小成本分析(CMA)等4种。在条件允许的情况下优先考虑成本-效用分析。

1. 成本效用分析(CUA)

适用于临床产出指标不同的抗肿瘤药品之间的比较分析。成本效用分析(CUA)通常将QALYs作为效用测量指标。使用该方法需要注意的问题是，不同的生命质量测量方

法、测量工具以及效用积分体系均会对效用值产生影响，因此需要具体阐述效用值测量及计算方法。

2. 成本效果分析（CEA）

成本效果分析（CEA）一般适用于比较药品只体现或主要体现在某一个临床产出指标时。基于药品临床综合评价数据来源，推荐使用医疗机构直接获取的临床指标，如 OS、PFS、ORR 等开展成本效果分析。

3. 成本效益分析（Cost Benefit Analysis, CBA）

CBA 的成本和健康产出均采用经济指标衡量，可将健康产出通过支付意愿法或人力资本法等方法转换为货币指标，进而进行净效益计算。在评估中需要解释健康产出转换成货币值采用的所有步骤和方法，并使用敏感性分析验证主要假设。当药品的健康产出难以采用质量调整生命年或其他任何单个效果指标测量时，可采用成本-效益分析。

4. 最小成本分析（Cost Minimization Analysis, CMA）

当有证据显示待评估的抗肿瘤药品及其对照药品的重要临床产出无统计学差异或无临床差异时，可选择最小成本分析的评估方法。在证明不同抗肿瘤药品临床产出的无差异时，统计学无差异和临床无差异均可以接受，当存在公认的临床无差异标准时，以临床无差异为准。

5. 增量分析

CUA 和 CEA 的基本决策原则是按照增量分析 (Incremental Analysis) 结果进行决策。增量分析是在干预方案与对照方案之间进行的成本和产出两个维度的比较,以增量成本效果比 (Incremental Cost-Effectiveness Ratio, ICER) 呈现结果。增量成本效果比是指评价药品与对照药品的相对成本和效果之差的比值。综合考虑效果与成本两个维度,如果 ICER 小于阈值,则评价药品成本效果优于对照药品,反之亦然。然而,目前我国还没有关于 ICER 阈值的统一标准,各国也没有统一的共识。不同疾病或健康状态下的阈值可能存在特异性。

为了验证研究结果的稳定性,建议对 ICER 结果开展敏感性分析,以便应对方法学及模型不确定性和参数不确定性。以上分析均可通过模型分析法实现,具体参见相关的药物经济学评价指南。

6. 不确定性分析

不确定性分析包括方法学不确定性、参数不确定性及模型不确定性等。研究者应当对经济学评价过程中的各种来源的不确定性进行全面分析。方法学及模型不确定性多采用情境分析。研究者应当明确定义不同情境下的分析方法和研究假设,并对不同情境分析的结果之间差异进行合理解释。参数的不确定性可以采用单因素、多因素分析法等确定型敏感

性分析，也可以采用蒙特卡洛模拟 (Monte Carlo Simulation) 进行概率敏感性分析。

参数敏感性分析用于研究系统对单个或多个输入因素变化的敏感程度：如药品价格、转移概率等，按因素取值确定与否可分为确定型敏感性分析和概率敏感性分析。参数敏感性分析范围需要有充分依据，来源于文献报告的 95% 置信区间、最大值和最小值，或者同类研究中参数估计的高值和低值。应尽可能纳入较多参数，参数分布等需给出合理说明。

单因素不确定性分析可用暴风图表示，明确各个不确定因素对结果影响的大小。需注意对暴风图坐标轴产出指标、分析变量等进行说明（见图 1）。建议使用成本效果可接受曲线或成本效果散点图(见图 2)展现概率敏感性分析结果。

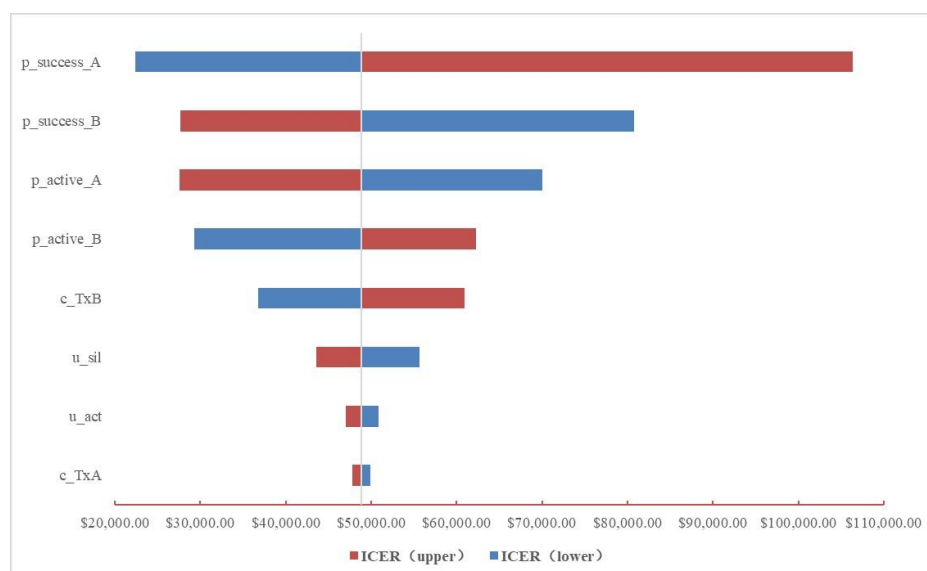


图 1 单因素分析暴风图示例

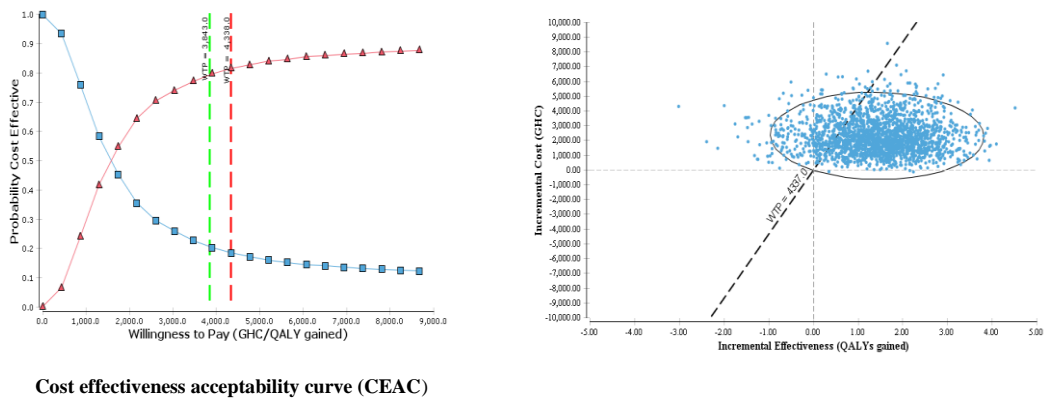


图 2 成本效果可接受曲线（左）与散点图示例（右）

7. 预算影响分析

可根据情况，选择是否针对特定药品的经济性开展预算影响分析。预算影响分析的实施可参考第四章相关内容。

四、创新性评估

抗肿瘤药品的创新性评估是一个专业性强、多角度、多层级的信息收集和研判过程，不同的评估角度和侧重点会有不同的结论。抗肿瘤药品的创新参与主体涉及政府、药商、经营企业、医生和患者，且对创新的理解和诉求各不相同。本着以患者为中心的原则，本指南从患者角度考虑药品的创新价值。

综合国内外文献结果和我国现实情况，建议从临床创新性和产业创新性进行评估，具体参考指标（见表 5）。临床创新性主要聚焦抗肿瘤药物对于患者用药需求的满足程度，指标包括：（1）满足临床尚未满足的诊疗需求，指在治愈或

改善目标疾病症状上，填补某一疾病治疗领域的空白，重点考虑对目标疾病的精准治疗，针对新的适应证有突破性的治疗效果，是否是国家特别关注、具有极高社会需求的产品类型，以引导药品创新发展方向等；（2）**用药创新**主要依据包括在疾病或伤痛治疗方面有更高的安全性、有效性和实用性；在治疗方案、适用人群、给药间隔、给药周期、药品剂型、给药途径、拆分包装和药品储存条件方面存在技术创新（如改善此领域技术短缺现状）等。

产业创新性关注国产创新及研发生产能力，考虑的核心指标包括：（1）是否为本土研发的原研药；（2）是否获得国内药品结构或制剂工艺方面的专利。

从上述两个创新性指标对所评估药品创新性程度得出差异性意见。例如：将创新性推荐等级分为三类，分别为强烈推荐、一般推荐和不推荐。此外，抗肿瘤药品的创新性评估亦可参照《化学药品注册分类改革工作方案》^[4]中对创新药的定义。创新性指标的测量可以通过专家问卷的形式实现量化，具体问卷设计、调查流程、计算方法可参见《问卷设计手册：市场研究、民意调查、社会调查、健康调查指南》^[5]。

表 5 创新性评估核心指标

分类	指标		数据来源	方法
临床创新性	存在未满足的临床需求	是否为肿瘤疫苗产品	药品批准信息；文献；专家咨询	文献研究；二手数据收集；定性访谈
		是否为罕见病肿瘤药/孤儿药		
		是否为儿童肿瘤用药		
	用药创新		文献；专家咨询；患者调查；	文献研究；定性访谈
产业创新性	是否为国内自主研发产品		文献；专利信息查找	文献研究；二手数据收集分析
	是否获得国内专利		专利信息查找；	二手数据收集分析
	是否为靶点创新		文献；临床专家	文献研究；二手数据收集分析；定性访谈
	是否为分子实体创新		文献；临床专家	文献研究；二手数据收集分析；定性访谈
	是否为制剂创新		文献；临床专家	文献研究；二手数据收集分析；定性访谈

五、适宜性评估

适宜性是衡量抗肿瘤药品临床使用情况的重要维度，适宜性评估是指运用药物流行病学等技术手段和统计方法对药品及其临床使用的可利用度进行科学评估的过程。

因抗肿瘤药品治疗疗程一般较长，药品适宜性对于提高患者的长期用药依从性具有重要作用。可从药品技术适宜性、药品使用适宜性、药品体系适宜性、药品监管适宜性四个层面进行评价。

（一）药品技术适宜性

药品技术适宜性可从药品标签标注、药品说明书、储存条件等方面进行评估。主要的指标应包括：标签标注完整性，药品说明书是否明确标注，起效快慢，剂型是否适宜于肿瘤患者服用，是否有特殊的存储条件，是否需要特殊装置，用药后是否需要监测或随访服务等。

（二）药品使用适宜性

药品使用适宜性评估应考虑以下维度：（1）给药适宜性，包括：药品用法是否符合肿瘤患者当时的身体状况，给药途径是否根据病人情况和药物特点来选择，给药方法的难易程度；（2）用药适宜性，包括：治疗方案、适用人群、给药间隔、给药周期等方面是否适宜。

（三）药品体系适宜性

药品体系适宜性主要关注在现阶段中国医疗政策，医疗、医药、医保三医联动的环境下。该药品对分级诊疗“基层首诊、双向转诊、急慢分治、上下联动”制度的影响情况。如该药品是否限定使用医院级别、限定门诊或住院患者使用、上下级医疗机构衔接情况、药师医生获得药品信息程度等。

（四）药品监管适宜性

药品监管的适宜性主要从卫生监管及医保监管两个方面进行考量。卫生的监管方面主要通过该药品在医院合理应用管理指标，如抗肿瘤药物分级管理制度执行情况；限制使用级和普通使用级抗肿瘤药物的使用率；抗肿瘤药物使用金额占比；抗肿瘤药物处方合理率与干预率；抗肿瘤药物不良反应报告数量及报告率；抗肿瘤药品临床应用监测及相关数据上报情况¹等。医保的监管主要看报销限定适应症、限定报销标准、是否存在超支付范围的使用等。

适宜性核心指标（见表 6）。适宜性评估信息主要通过文献信息收集和问卷调查实现。在对药品适宜性进行具体评估时，大多指标为定性的证据资料，可对药品不同角度适宜性指标进行程度的划分后，根据可获得相关适宜性材料对该指标适宜性优劣的选择，但需在报告中详细阐明划分的标准及依据。此外，也可对适宜性指标定性资料的完整性、证据来源的质量进行描述性分析。

资料来源主要包括：药品相关政府网站信息，如：国家药品监督管理局（National Medical Products Administration, NMPA）网站、美国食品药品监督管理局（Food and Drug Administration, FDA）网站等；药品相关材料，如：药品说

¹ 国家卫生健康委关于印发抗肿瘤药物临床应用管理办法（试行）的通知.国卫医函【2020】487号
<http://www.nhc.gov.cn/yzygj/s7659/202012/a7600740bed44d1db7015ca5a1be2cc0.shtml>

说明书、药品注册资料、媒体广告、市场调查资料（供应情况）等；文献数据库，如：中国医院知识资源总库（China Hospital Knowledge Database, CHKD）和企业申报资料等。也可通过医务人员调查问卷（内容包括药物的起效快慢、药品使用方法的难易）收集信息。适宜性指标的测量可以通过专家问卷的形式实现量化，具体问卷设计、调查流程、计算方法可参见《问卷设计手册：市场研究、民意调查、社会调查、健康调查指南》^[6]。

表 6 适宜性评估核心指标

分类	指标	数据来源	方法
药品技术适宜性	药品标签标注的完整性	说明书；监管部门网站；生产企业提供信息	查阅药品监管部门网站说明书；访谈
	药品说明书是否明确标注剂量标准	说明书；临床指南；监管部门网站	查阅药品说明书、临床指南、监管部门网站信息；访谈
	是否有特殊的存储条件	说明书；生产企业提供信息	查阅药品说明书、临床指南、监管部门网站信息；访谈
	是否需要特殊装置	说明书；生产企业提供信息；调查问卷	查阅药品说明书、临床指南、监管部门网站信息；访谈
	用药后是否需要监测不良反应	说明书；生产企业提供信息；调查问卷	查阅药品说明书、临床指南、监管部门网站信息；访谈
	药品起效快慢	说明书；问卷调查	查阅药品说明书、临床指南、监管部门网站信息；访谈
药品使用适宜性	给药途径是否根据患者情况和药物特点来选择	调查问卷；专家咨询	文献分析、访谈
	药品使用疗程长短情况	调查问卷；专家咨询	文献分析、访谈
	给药方法难易程度	调查问卷；说明书	文献分析、访谈
	患者服药时间间隔长短情况	调查问卷；专家咨询	文献分析、访谈
	药品剂型是否方便医生、药师、患者使用	调查问卷；专家咨询	文献分析、访谈

分类	指标	数据来源	方法
药品体系适宜性	是否限定使用医院级别	监管部门网站	查阅监管部门网站信息
	是否限定门诊或住院用药	监管部门网站	查阅监管部门网站信息
	该药在上下级医疗机构衔接情况	调查问卷；专家咨询	文献分析、访谈
	药师医生获得药品信息程度	调查问卷；文献数据库	文献分析、访谈
药品监管适宜性	合理用药情况	调查问卷；专家咨询	文献分析、访谈
	医保报销适应症	调查问卷；监管部门网站	文献分析、访谈、查阅监管部门网站信息
	医保报销比例	监管部门网站	查阅监管部门网站信息
	是否存在超支付范围使用的情况	调查问卷；监管部门网站	查阅监管部门网站信息；访谈

六、可及性评估

参考 WHO/HAI 药物可及性标准化方法，抗肿瘤药品的可及性评价主要涉及可获得性和可负担性两个方面。

（一）可获得性

可获得性可界定为肿瘤患者获得目标药品潜在机会的大小，可由医疗机构药品配备使用情况或有无短缺情况等反映。此外，药品生产企业的生产能力、药品在医疗机构和零售药店的分布、配送公司配送能力等会影响药品的可获得性，可根据评估需要可从不同渠道获得相关支持信息。可获得性评估的主要指标（见表 7）。

表 7 可获得性评价的参考指标

评价内容	指标	数据来源	方法
医疗机构药品配备使用情况	药品配备率	行业协会报告，国家临床用品使用监测；调查	通过针对综合医院、基层医疗卫生机构的分层抽样调查获得

评价内容	指标	数据来源	方法
药品短缺情况	药品是否存在短缺	全国短缺药品信息直报系统；多源信息采集平台	二手数据收集与分析；访谈；问卷
药品生产能力	药品产能	生产企业提供信息；第三方数据	二手数据收集与分析；访谈；问卷
医疗机构覆盖情况	药品对于各级医疗机构的覆盖数量	生产企业提供信息；第三方数据	二手数据收集与分析；访谈；问卷
	所覆盖医疗机构的地理分布	生产企业提供信息；第三方数据	二手数据收集与分析；访谈；问卷
零售药店覆盖情况	药品对于各级行政区域的覆盖数量	生产企业提供信息；第三方数据	二手数据收集与分析；访谈；问卷
	所覆盖行政区域的地理分布	生产企业提供信息；第三方数据	二手数据收集与分析；访谈；问卷
配送能力	配送商业合作商的数量	生产企业提供信息；第三方数据	二手数据收集与分析；访谈；问卷
	配送商业合作商的地理分布	生产企业提供信息；第三方数据	二手数据收集与分析；访谈；问卷

(二) 可负担性

可负担性由药品费用与收入水平的关系来反映，可从疗程费用负担和年费用负担两个方面考虑：（1）对于疗程的费用负担，计算每疗程的药品费用与最低日薪标准的比值，若药品费用低于最低日薪标准 1 倍，认为该种药品具有良好的可负担性。急性病的疗程一般为 7 天，慢性病的疗程一般为 30 天，可根据疾病的实际诊疗情况进行调整并说明。（2）对于年费用负担，计算人均药品年费用占家庭年可支配收入的比重（%）。一般认为该比重大于等于 40% 时，即造成灾难性医疗支出，该种药品不具有良好的可负担性。必要时应

了解医保报销情况，计算经社会医疗保险和商业医疗保险保障后患者的支付水平，以判断药品的实际可负担性。由于我国城乡居民收入水平差异明显，可分别针对城镇和农村居民进行药品可负担性评价，考虑不同的收入水平下的药品可负担性。可负担性评价的数据来源（见表8）。

表8 可负担性评价数据来源参考

评价内容	相关数据	数据来源	方法
疗程费用负担	药品价格	国家或省级药品集中采购平台、药品信息数据库、线上药店	查阅国家或省级药品集中采购平台、药品信息数据库、线上药店数据
	疗程	药品说明书；国内诊疗指南/专家共识；WHO示范处方集	查阅药品说明书；国内诊疗指南/专家共识；WHO示范处方集
	最低工资标准	国家人力资源和社会保障部网站	查阅国家人力资源和社会保障部网站
年费用负担	药品价格	国家或省级药品集中采购平台、药品信息数据库、线上药店	查阅国家或省级药品集中采购平台、药品信息数据库、线上药店
	家庭年可支配收入	国家统计局网站	查阅国家统计局网站

第二章 评价设计

一、评价技术类型

根据研究时限、数据可得性和研究目的，可将抗肿瘤药品临床综合评价分为快速综合评价和完整综合评价两种类型（见图 3）。研究者可根据具体情况选择合适的评价类型。但无论哪种评价类型，其评价方法都可能包含文献分析、试验设计、定性访谈、模型研究等四个方面。

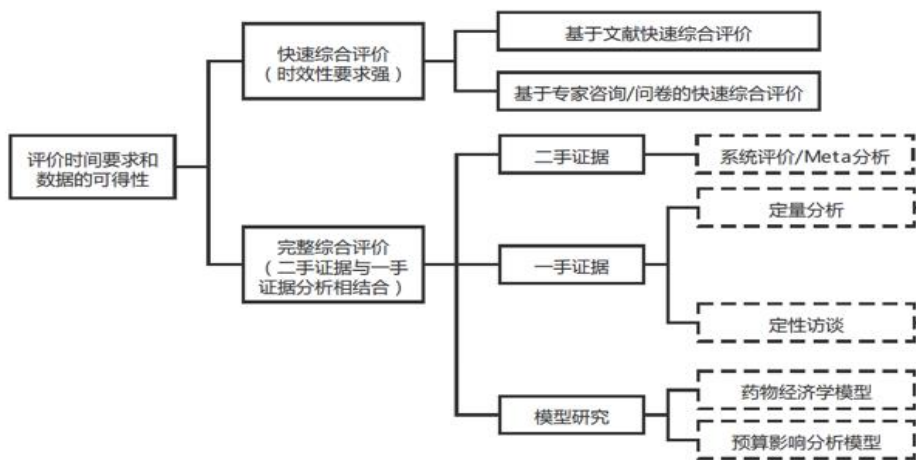


图 3 评价类型及基本设计

（一）快速综合评价

当研究时限紧迫或二手证据资料较为充分时，通常采用快速综合评价法评价目标药物^[3, 7, 8]。快速综合评价是通过检索 1-2 个数据库提取证据等级相对较高的证据或者现有资料

来对干预技术进行综合的评判。

1. 基于文献快速综合评价

根据评价问题、研究人员数量及时间要求，可分为系统性文献综述（systematic review）和描述性文献综述（narrative review）两种，前者比后者需要的人财物资源更大，适合在文献资源较多、证据等级较高的条件下实施。系统性文献综述利用循证医学及证据质量评估等方法工具，对国内外临床研究文献证据进行综合评估，基于证据质量及推荐强弱程度，报告该干预技术特性及其临床安全性、有效性或经济性文献评估结果。推荐使用 Cochrane 协作网有关系统评价方法及工具，必要时应对临床研究文献证据进行 Meta 分析。

描述性文献综述适用于目的较为明确的信息检索，如检索政府、行业协会及国际组织网站查找特定证据或信息。

2. 基于专家咨询/问卷的快速综合评价

针对不同品牌、规格、剂型的同类药品的特点分析（可及性、合理用药等），可采取基于专家咨询或问卷调研的快速综合评价，集合有关信息，进行具体药品间比较分析，支持快速综合评价的开展。

（二）完整综合评价

当研究时限允许时，可选择完整综合评价法评价目标药品，提高研究结果的可信度。完整综合评价^[9, 10]通常包含临

床试验设计，建议根据流行病学、卫生统计学、卫生技术评估的标准评价流程及方法进行设计。建议在必要时采用多中心临床研究方法，在更大的范围内收集病例资料，提升抗肿瘤药品临床综合评价结果的代表性和可信度，保证研究质量。完整综合评价维度^[8]应包含其临床效果、安全性、费用以及经济性影响、社会与伦理、文化与法律、组织架构、医疗体系以及利益相关者。

针对安全有效性评估，应采取二手证据（文献分析等）与一手证据（临床试验研究、定性访谈等）分析相结合方式开展，基于二手证据评估结果，确定一手证据评估需求，定义评估问题及关键指标，建立基于临床一手数据分析的评估方案；针对经济性评估，建议结合临床现实数据进行成本（费用）等分析，必要时可根据文献参数等建立远期成本效果分析模型或预算影响分析模型，可参考 2020 版中国药物经济学评价指南。具体流行病学和卫生统计学方法可参考全国高等学校统编教材、卫生技术评估方法可参考卫生技术评估教材。

二、评价流程

通过主题遴选确定评估主题，在完成项目委托的基础上，开展具体评估过程，主要包括评估实施、证据审核、公开质询和证据转化四个步骤（见图 4）。

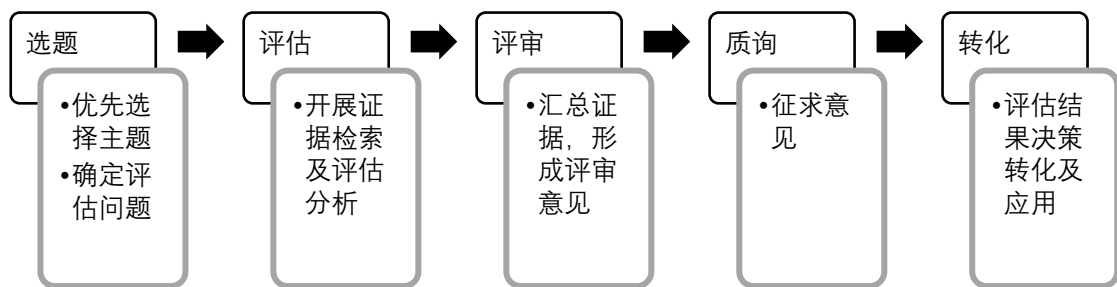


图 4 药品临床综合评价流程

三、抗肿瘤药品临床综合评价实施路径

(一) 评价设计

1. 评价原则与步骤

遵循 PICO 原则，即通过明确目标人群（P），明确干预措施（I）和对照措施（C），确定结局指标（O）等内容，对评价主题进行科学、合理设计。

评价步骤包括确定拟解决的评价问题（评价背景和评价目的）、评价角度、药物治疗的干预和对照、研究人群、评价的维度和测量指标、评价方法和数据来源，以及评价计划和时间安排。

2. 明确问题

评价临床和成本效益之前应清楚地描述评价问题，明确干预措施和相关患者人群，评价问题应与评价范围相一致。描述评价问题时应包括所有主要的相关技术及其在临床路径中的位置。

3. 评价范围

评价范围决定证据评价的范围和内容，为评价过程提供指导框架。基于 PICO 原则明确评估范围。当需要做出推荐意见时可采用 PICOS，S 为 setting 在此代指特定的环境/背景，具体包括：

(1) 评价背景：描述与技术相关的疾病、健康状况，包括疾病的预后、流行病学以及经济负担，主要干预措施（包括药物与非药物）及其疗效于安全性，国内外临床诊疗指南对治疗方案的推荐，全球范围内相关干预措施的药物经济学评价现状（基本结论和尚存的问题），以及本研究的价值（必要性和重要性）等。

(2) 评价目的：明确提出本次抗肿瘤药品临床综合评价的主要目的和待解决的问题。评价目的中应简明扼要地阐述“运用何种理论和方法，解决何种主要问题，达到何种主要目的”。评价目的要与评价背景所阐述的问题相互呼应。

(3) 评价角度：研究者应根据研究目的和报告对象明确评价角度。评价角度主要包括以下几类：全社会角度、卫生体系角度、医疗保障支付方角度、医疗机构角度以及患者角度。

当研究目标服务于卫生政策决策时，推荐采用卫生体系角度进行评价；当研究属于纯理论或方法学研究时，可考虑全

社会角度的评价；对于其它情形，研究者可根据评价目的选择合适的评价角度。

（4）适用（目标）人群：评价需要明确待评药品（干预）的目标人群以及纳入标准与排除标准。一般情况下，抗肿瘤药品临床综合评价的目标人群应与药物的适应症人群保持一致。

应采用国际疾病分类编码（如 ICD-10）界定适应症。当目标人群存在较大的异质性时，可以根据研究需要开展亚组分析，如根据人口特征、疾病亚型、严重程度和合并症进行分层分析。

（5）待评药品（干预）与对照药品：待评药品和对照药品的描述应包括剂型、规格、用法用量、给药方式、合并用药和治疗路径等信息。待评药品和对照药品应以通用名表示，同时列出商品名。

对照药品的选择建议依据最新临床指南，选取指南推荐的该适应症标准治疗方案或常规治疗方案。若没有标准治疗方案，可依据专家共识或临床专家意见，纳入临床常规治疗方案。临床专家的选取需要考虑专家数量及专家代表性。若对照药品不符合上述标准，可考虑与安慰剂进行比较，但须说明其使用安慰剂比较的临床合理性。根据评价目的，可选择一个或若干个对照药品。

(6) 评价维度及指标选择:评价维度是评价设计的核心内容。常用评价维度包括:安全性、有效性、经济性、创新性、适宜性、可及性及药物政策评价维度。应针对决策问题,选择确定具体纳入评价的维度。纳入或不纳入某个具体维度,均应在综合评价的设计方案中做出明确说明。在各维度内确定相应的测量和评价的指标,具体要求参见“第二章”的相关内容。

(7) 研究方法选择:抗肿瘤药品临床综合评价采用定性定量结合的方法,收集一手和二手资料进行分析。对安全性、有效性和经济性进行评价时,首先采取系统文献综述收集和/或二手资料分析,若仍未满足评价的需求,应增加一手资料的研究。进行一手资料收集时,应首选基于医疗机构临床数据的真实世界数据研究。历史数据若未能满足需求时,可开展前瞻的观察性研究或干预性研究(临床试验研究)。

(二) 证据整合

证据的整合与评价将在主题以及评价维度确定后实施。我们需要全面的收集对于每一个待评估技术的所有的评价维度信息,并确保纳入多来源的数据用于支撑评价体系。同时,评价实施主体需要对证据的质量予以界定。

（三）评审决策

证据评审是将卫生技术评估项目产生的结果及其他信息进行整合评估，从临床价值、经济价值、创新价值和社会价值等维度判断一项卫生技术的综合价值，并做出最终决策意见的过程。

证据评审由评审专家组负责，评审专家组基于可得证据，在充分咨询医药生产及经销企业、医患代表等各方意见后，做出关于特定技术临床合理应用、应用推广等决策。

（四）结果征求意见

为确保评估评审结果代表各方意见，评价工作实施主题应对评估报告扩展版摘要及评审决策进行公开质询以确保其质量。质询旨在通过公开咨询的方式以保障评估评审结果的公正性和合法性，使更多关注技术的利益相关方及公民社会参与技术的决策过程，从而提高技术决策的科学性。质询对象为专家、患者以及其他评估实施主体。评审意见产生后，评估实施主体将整理评估报告扩展版摘要，连同评审意见，一并发布在官方网站上，全面征求有关各方意见。征求意见期限为有关资料挂网后的 15 天内。

(五) 应用转化

基于专家评审结果，可将研究成果进行相应的结果转移转化，如医院内部用药调整等。

第三章 评价方法

一、基于一手资料的评价方法

（一）真实世界数据

药品临床综合评价应围绕基本用药决策需求，结合临床现实，规范、科学、合理地设计并实施临床研究，推动真实世界数据（Real World Data, RWD）在临床综合评价中的使用并发挥其优势。参照药品审评中心针对真实世界研究相关实效性试验研究设计、观察性研究设计和其他非试验设计的推荐意见，定义疾病组别与药品特性密切相关的真实世界数据来源及分类标准，编制规范的药品临床综合评价数据集，定义采集范围、采集变量、采集方式等。充分依靠医院信息系统等现有电子信息系统采集现有数据，确保数据采集的准确性、真实性和完整性。

鼓励医疗机构等药品临床综合评价主体建立相关审查监督制度，对数据获取、数据质量、分析过程、结果阐释等关键环节进行质控评估，强化科研伦理管理及患者隐私保护，避免数据收集及分析偏倚，有效地支持药品临床应用安全性、有效性及经济性评价开展。

1. 真实世界数据定义

药品临床综合评价应充分利用真实世界数据。真实世界数据是指来源于日常所收集的各种与患者健康状况和/或诊疗及保健有关的数据。并非所有的真实世界数据经分析后能成为真实世界证据，只有满足适用性的真实世界数据才有可能产生真实世界证据。

2. 真实世界数据来源

当前我国药品临床综合评价相关的现有真实世界数据来源主要包括但不限于：

(1) 医院信息系统数据：医院信息系统数据包括结构化和非结构化的数字化或非数字化患者记录，如患者的人口学特征、临床特征、诊断、治疗、实验室检查、影像学检查、生化检验、基因检测、安全性和临床结局等。

(2) 医保支付数据：医保支付数据主要来源于政府、医疗机构建立的基本医疗保险体系，包括有关患者基本信息、诊断、医疗服务利用、处方、结算年度、费用、报销金额等结构化字段的数据。

(3) 登记研究数据：登记研究数据是通过有组织的系统，利用观察性研究的方法搜集临床和其它来源的数据。登记研究数据库一般包括患者自报数据和长期随访数据，观测结局指标丰富且具有准确性较高、结构化强等优点，对于评价药品的安全性、有效性、经济性和依从性具有较好的适用性。

(4) 药品监测系统: 药品监测系统主要用于开展药物安全性、适宜性、可及性研究。包括国家或区域的药品使用监测系统、短缺药品监测预警系统、药品不良反应监测系统等。

(5) 自然人群队列数据: 自然人群队列数据指对健康人群和/或患者人群通过长期前瞻性动态追踪观察, 获取的各种数据。此类真实世界数据可以帮助构建常见疾病风险模型。

(6) 死亡登记数据: 人口死亡登记数据包含死亡医学证明书中的所有信息, 包括死亡原因、死亡时间等, 此类数据可作为重大疾病临床结局的数据来源。

(7) 患者报告结局数据: 患者报告结局是一种来自患者自身测量与评价疾病结局的指标, 包括症状、生理、心理、生命质量等, 患者报告结局在药物评价体系发展中越来越重要。

(8) 政府网站数据: 包括国家药品监督管理局(NMPA)、美国食品药品监督管理局(FDA)、国家药品不良反应监测中心等政府网站、国家统计局。

使用真实世界数据开展药品临床综合评价前, 应对数据适用性进行充分评估, 围绕真实世界证据可以回答的临床与卫生健康政策问题, 进行科学的研究设计和严谨的组织实施, 获取相关、可靠、适宜的真实世界数据, 进行恰当、充分、准确的分析后, 可形成药品临床应用安全性、有效性、经济性等相关证据。

（二）真实世界研究的基本步骤

真实世界研究与传统临床研究过程总体上而言是类似的，仍然需要经过确立研究问题，建立研究方案，获取数据形成研究数据集，分析处理数据和报告等几个过程（图 5）。最主要的区别在于：在 RWD 的研究环境下，具体的研究问题可能是在已经建立了研究型数据库之后产生的。因此，对于数据库和数据库框架体系的了解和掌握，很大程度上决定了研究的可行性和科学性。

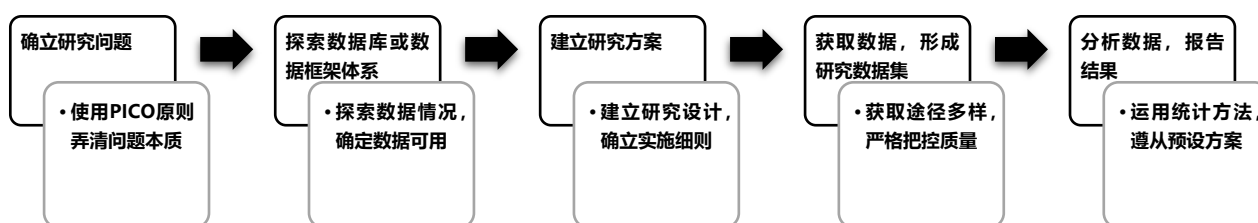


图 5 真实世界研究的基本步骤

（三）偏倚的控制方法

在使用真实世界数据开展药品临床综合评价研究中，大多数为基于常规收集的医疗数据的观察性研究，相比随机化临床试验，这类研究更容易产生偏倚，常见的偏倚类型包括选择偏倚（Selection Bias）、信息偏倚（Information Bias）和混杂偏倚（Confounding Bias）。利用统计学方法尽可能减小偏倚对研究结果的影响，是使用 RWD 开展药品临床综合评价研究的关键环节。

1. 选择偏倚

选择偏倚是指选入分析的研究对象与未选入者在某些特征上存在差异而引起的系统误差。入院率偏倚（研究对象选择某个医院的某类疾病患者，而非全体目标人群的一个随机样本）是临床综合评价研究中常见的选择偏倚。在统计分析中需要提供以下信息，来评估研究人群是否存在一定的选择偏倚，以保证研究结果的外推性：①研究报告中提供研究对象的筛选流程图，需说明研究过程中每一个步骤剔除的样本量以及相应的剔除原因。②研究报告中提供研究对象的基线特征（例如针对肿瘤患者，基线特征包括年龄、性别、种族、肿瘤类型、疾病分期、既往治疗、ECOG 评分等），可按干预方案分组详细描述患者的基线特征并进行统计学检验。③研究报告中提供纳入患者和排除患者的基线特征对比情况并进行统计学检验。

2. 信息偏倚

信息偏倚是指在收集整理信息过程中由于测量暴露或结局的方法有缺陷造成的系统误差。异常值现象是临床综合评价研究中常见的信息偏倚。本指南推荐在研究中应对重要的变量进行异常值检测，确定是否存在异常值。如存在异常值，推荐与临床医生共同判断该数值真实存在的可能性，而不推荐任意删除或修改异常值，导致该变量原始分布的变动。

3. 混杂偏倚

混杂偏倚是指在数据分析阶段，由于某些非处理因素与试验因素对效应指标的共同作用，使统计分析结果产生偏倚，从而影响结论。为减少混在偏倚对于研究结果的干扰，本指南推荐在统计分析计划中预先制定和选取如下的统计分析方法：①倾向性评分，倾向性评分（propensity score）是指在一定协变量条件下，一个观察对象可能接受某种处理（或暴露）因素的可能性。其基本原理是将多个混杂因素的影响用一个综合的评分来表示，从而减少了自变量的个数，有效的克服分层和匹配中要求自变量个数不能太多的短板。倾向性评分调整混杂的方法主要分为四大类，分别为倾向性评分匹配法，逆概率加权法，倾向性评分分层法和倾向性评分作为协变量调整法。在使用倾向性评分方法控制混杂因素时，本指南推荐优先使用倾向性评分匹配法，但当对照组和治疗组倾向性评分分布重合较小、匹配后样本损失多、代表性大大降低时，推荐使用逆概率加权法。当以上两种方法均不适用时，可考虑使用倾向性评分分层法或倾向性评分作为协变量调整法。②多变量分析，多变量分析方法是在临床综合评价研究中较常见的控制混杂因素的统计分析方法。研究中应根据研究目的、研究设计类型、治疗方案的特点、结局变量及混杂因素选择合适的统计模型。统计模型详见表 9。

表 9 多变量回归模型^[11]

结局变量	单次测量		重复测量
	独立样本	聚集性样本 (如多中心研究)	
二分类结局变量	Logistic 回归	多水平 Logistic 回归、 GLMM、GEE、条件 Logistic 回归	GLMM、 GEE
连续性结局变量	线性回归	多水平线性回归、 GLMM、GEE	GLMM、 GEE
生存时间	Cox 比例风险回 归、竞争风险模 型	方差-校正 Cox 模型或共 享脆弱模型	
累积或计数数据	泊松回归	多水平泊松回归	

注：GLMM：广义线性混合模型；GEE：广义估计方差。

在研究中使用这些模型时，应考虑以下注意事项：①确保模型的基本假设成立，如 Cox 比例风险模型的等比例风险假设；②确保变量结构与分布（如正态性）满足模型的要求；③纳入的研究对象数量足够满足模型中需要估计的参数所需的例数，如一般研究对象的数量至少为 20~30 倍的协变量数，发生结局事件的患者数量建议至少为 10 倍的协变量数；④考虑模型中的重要因素与结局事件是否存在非线性关系。

4. 未知混杂因素

以上方法仅能控制已测量到的混杂因素所引起的偏倚，而无法消除未知的、被遗漏的、以及测量不准确的混杂因素所造成的残余混杂。为进一步消除这部分残余混杂，本指南

推荐在使用真实世界数据开展临床综合评价研究时，应尝试使用双重差分法（Difference In Difference, DID）、工具变量法（Instrumental Variable, IV）等方法控制未知的混杂因素。

（四）真实世界药物经济学评价

药物经济学评价模型通常是采用图形结构、公式等方式对疾病的自然转归过程和干预措施对该疾病转归过程的影响进行抽象模拟，重点关注此过程中发生的干预措施和重要临床事件，以及由此引起的健康变化和资源消耗情况，最终在不同干预方案之间进行经济性比较。在抗肿瘤药品经济学评价中，比较常用的经济学评价模型是马尔科夫模型和分区生存模型，模型分析的相关介绍可参考《中国药物经济学评价指南（2020版）》，本指南将重点阐述模型中各类参数对应的真实世界数据来源。

1. 安全性数据

药品不良反应数据可以来自于药品安全性主动监测数据或医疗机构药品不良反应上报信息。与药品不良反应相关的预防、监测或治疗成本，需要纳入模型中进行计算。

2. 有效性数据

抗肿瘤药品有效性评价指标详见本指南第二章，主要临床结局指标为经济学评价中重点关注的有效性指标。主要临床结局指标包括总生存期和无进展生存期，总生存期可结合

医院信息系统中患者首次处方目标药品时间和死亡登记数据中患者死亡时间、死亡原因等信息，通过患者唯一识别字段进行链接获取。无进展生存期可通过医院信息系统中患者目标药品处方信息结合实验室信息系统和影像信息系统中记录的病理、影像等信息获取。

对于安全性以及有效性维度真实世界数据证据等级评定，应依据选取的研究设计与研究问题之间的相关性、研究质量控制程度、以及选取的研究数据的可靠性综合评定。本指南借鉴《真实世界研究指南》中对于常见的真实世界研究类型的证据等级由高到底排序，依次为实用性临床研究、前瞻性队列研究、回顾性队列研究、病例对照研究、横断面研究、单病例研究（见图6）。

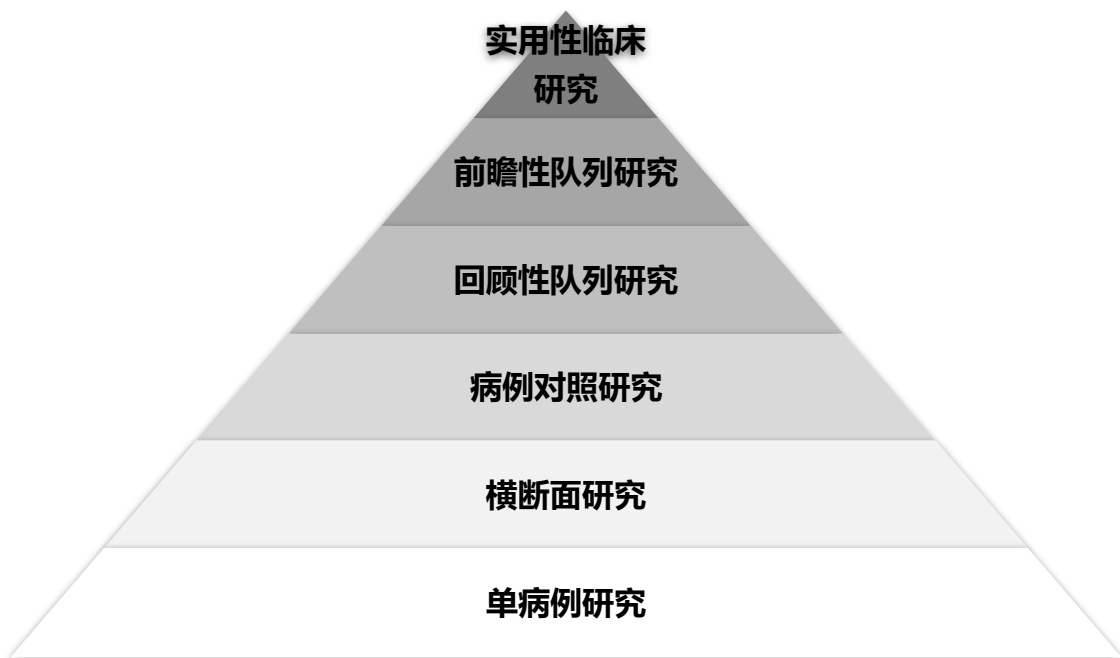


图6 真实世界研究的证据等级

3. 成本数据

成本数据可按照各个项目总的费用和明细费用分为两部分。其中总的费用包括患者一次就诊总费用、检查费、床位费、诊查费、治疗费、化验费、检验费、耗材费等；明细费用包括患者所接受的综合医疗服务、诊断、治疗、药品、康复、耗材等项目下具体细项目的类别名称、项目的名称、数量和单价数据。成本数据可来自于医院信息系统中病案首页数据、医保支付数据和登记研究数据等。

4. 健康效用值

健康效用值可来自于患者报告结局测量数据或者相关研究文献。非效用测量量表测量的结果不可以通过百分制转化方法转换成“效用值”，应采用映射法（Mapping）或其他方法进行转换获得。

5. 预算影响分析

研究者可依据研究目的选择是否有必要开展预算影响分析。预算影响分析（BIA）旨在判断某个抗肿瘤药品入市、临床使用或扩大应用对于医院、医保及卫生财政的经济影响。预算影响分析测算时间区间一般为 3-5 年。

预算影响分析基本流程包括：

1. 结合临床实际用药情况，建立预算影响分析模型；
2. 通过文献检索、试点机构调查和医护人员等知情人访谈等方式，建立参数表。主要的参数应包括：疾病负担、患者治疗情况、收费报销等背景参数；

3. 模拟药品替代或扩大应用后的不同情景，分析对于公共预算、医院运行成本、医保支出、患者费用负担的影响。

选择预算影响分析的数据源时，应当从决策者视角出发。本指南推荐优先选取真实世界数据。与决策者视角相符的真实世界研究数据，可以为预算影响分析提供接近真实情况的使用量与成本信息，这些研究可以是注册登记研究也可以是基于数据库的回顾性研究。

（五）数据信息安全

坚持“谁主管谁负责、谁授权谁负责、谁使用谁负责”原则，加强评价过程中的数据收集、存储、使用、加工、传输、提供、公开等环节的安全管理。各评价任务承担机构应当建立健全相关信息网络安全管理制度、操作规程和技术规范，严格执行患者隐私保护和国家保密规定，构建可信的网络安全环境。任何单位和个人不得非法获取或泄露数据，未经国家及省级组织管理部门授权，不得擅自使用或发布国家及省级药品临床综合评价相关数据信息。各评价实施机构和人员对其组织实施评价工作任务范围内的数据、网络安全、个人信息保护和证据质量承担主体责任。

（六）专家咨询法

1. 定量咨询数据

当真实世界数据、文献资料等来源都不能满足临床综合评价数据需求时，可考虑采用专家咨询法。专家咨询法一般不可以用来获得临床治疗的有效率、不良反应发生率等关键参数的取值。如果研究中的某些参数采用了专家咨询法获取取值，则需要明确说明专家咨询的问题大纲、抽样方法、调研人数、调研方式、调研数据处理方法等细节。对通过专家咨询法得到的定量参数，建议进行敏感性分析。

2. 定性访谈数据

定性访谈，尤其针对专家的访谈，是临床综合评价中的重要方法，在各维度研究中均可涉及。定性访谈通过收集利益相关方或专家对具体问题的看法、认知及态度，对所研究的主题进行定性分析及判断。定性访谈方法主要分为以下两类：关键知情人访谈和一般咨询会议。

(1) 关键知情人访谈

关键知情人访谈是常用的定性访谈方法，一般需要确定访谈对象，并根据对象特点设计半结构化访谈提纲，访谈过程中做访谈记录，并整理访谈内容，形成文字材料，从而获得定性数据，具体内容如下第一步通过文献检索以及预设定的评估主题，确定访谈内容与指定访谈提纲，并就问题选择适宜的访谈对象，可选择各级卫生、药监、医保等政府部门有关负责人，医疗机构负责人、医护人员、药品企业和患者代表以及相关领域专家等。第二步通过抽

样或筛选的方式逐层选择上述可选择访谈对象，并就对待评估药品以及对照药品的临床使用安全性、有效性、经济性、创新性、适宜性和可及性的看法、认知及感受等信息进行深入的文本信息发掘；第三步根据访谈内容，进行文本资料的整理与分析。

(2) 一般咨询会议

一般咨询会议是定性获得专家意见的常用方式。常适用于对临床综合评价方案设计的讨论。以面对面座谈的方式，主要咨询行业内权威的专家，对方案设计予以评价与关键节点的把控，主要流程为：第一步建立主题，主题定义清晰，涉及问题方面不宜太多，应尽符合专家背景和水平范围，确保专家理解待咨询内容的角度趋同；第二步组织与会专家发表意见并进行充分讨论；第三步就讨论问题达成共识，取得较一致的结论。

二、基于二手资料的评价方法

通过文献综述法，整合已公开发表经过同行评审的文献、灰色文献、卫生医保药监等政府机构的官方网站资料等。基于文献综述总结的证据可用于支撑药品在安全性、有效性、经济性、创新性、适宜性、可及性等维度下的不同指标。

文献综述可分为描述性综述（Narrative Review）、系统文献综述和荟萃分析（Systematic Review and meta-analysis）

和其他文献综述（Review of Reviews, Scoping Review）。推荐开展系统文献综述，整合药品的临床安全性、有效性、经济性证据；但由于系统文献综述花费时间较长，也可视情况考虑开展描述性综述。推荐开展描述性综述，整合药品的创新性、适宜性、可及性证据。系统文献综述和描述性综述的差异（见表 10）。开展描述性综述时，推荐在质量评价与证据整合之前也参照系统文献综述的做法，尽量减少偏倚。

表 10 描述性综述与系统文献综述的差异

特点	描述性综述	系统文献综述
研究问题	宽泛或不存在	具体
文献检索	细节不明确	严谨、全面
文献筛选	潜在偏倚	严谨、基于标准
质量评价	变异度较大	严谨
证据整合	不常见	常见

系统文献综述被认为是循证医学证据等级最高的研究证据。根据 PICO 原则开展系统综述，其操作步骤主要为：

（1）确定研究问题；（2）文献检索；（3）应用纳入和排除标准；（4）数据提取；（5）数据分析。

（一）确定研究问题

常用的临床问题可分为：诊断、病因、治疗、预后和预防等类型。在文献检索之前，通常需要将复杂的研究问题转换成易于开展研究的结构化研究问题。通常情况下，推荐将问题转化成 PICO (TS) 形式，有助于研究的开展。P

(Population) 代表研究目标人群，一般为适应证人群；I (Intervention) 代表待评价的干预药品，通常为临床药品综合评价的研究对象；C (Comparator) 代表对照组，通常为当前该适应证的最佳治疗方案；O (Outcome) 代表效果指标，效果可以为临床效果或经济学评价结果。临床效果推荐使用主要结局指标 (Primary Outcome)，如总生存期 (OS)，也可使用次要结局指标 (Secondary Outcome) 或替代指标 (Surrogate Outcome)。经济学评价结果推荐使用增量成本效果比 (ICER)，但需严格阐述模型角度、框架、成本计算过程等关键信息。必要时，也可给研究问题增加研究时间 (T, time) 和研究类型 (S, study) 或研究背景 (S, setting)，使研究问题更加明确具体。

(二) 检索文献

根据前期对研究问题的整理归纳，形成简练、有效的检索主题词和自由词并确定相关检索源，随后制定有效的检索策略，确保检索的全面性。建议检索中英文数据库，以确保纳入文献的广泛性。常用临床证据的英文数据库如下：MEDLINE、EMBASE 和 Cochrane 中心对照试验注册库 (CENTRAL) 常用中文数据库为：CNKI、万方、维普和中国生物医学文献数据库 (CBM) 等。此外，根据研究需要，推荐纳入如下几类数据库：(1) 其他专业领域数据库，如 PsycINFO、

AMED、CINAHL；（2）未发表的在研研究：ClinicalTrials.gov 和 WHO 建立的国际临床试验注册平台^[12]（ICTRP）。常用的经济学评价数据库有 CDRs NHS Economic evaluation database（NHS EED）、The Health Economic Evaluation Database（HEED）和卫生经济学数据库可参考英国约克大学评审与传播中心（Centre for Reviews and Dissemination, CRD）官网（<https://www.crd.york.ac.uk/CRDWeb/>）等。灰色文献作为重要的证据来源，也推荐纳入到检索范围，建议参考加拿大国家卫生技术评估中心（CADTH）发表的“Grey Matters”^[13]查询文献来源。可参考麦克马斯特大学的“hedges”^[14]制定检索策略，检索综合性数据库。

建议关注证据级别高和推荐意见强的报告，如 GRADE 系统推荐的高质量证据。对文献质量的评定，也可参阅 Cochrane 文献质量评价标准或与相关 Cochrane 评价小组或相关检索专家联系。

（三）建立纳入、排除标准

根据综述关注的研究领域、目标人群、评价药品、结局指标、研究类型等相关因素，建立合理的纳入和排除标准，对文献进行筛选。同时按照题目/摘要初筛、全文筛选和证据等级评价的排序，对检索到的文献资源进行筛查。

(四) 证据评价与证据整合

基于系统评价的证据质量进行评价，包括两个方面的内容，一是对纳入系统评价的单个研究的偏倚风险评估，二是对总体证据的质量分级。偏倚风险评估因针对不同临床问题如治疗、病因、诊断和预后的系统评价所纳入的单个原始研究的设计类型和实施方式并不相同，其质量评价工具和方法也不尽相同，具体方法参照 Cochrane 干预措施评价手册。对总体证据的质量评价和分级，本指南推荐使用 2004 年 GRADE 工作组制定的证据质量和推荐强度分级系统 (GRADE) 对证据质量进行评价^[15]。根据证据的同质性与异质性，可使用随机效应模型或固定效应模型对证据进行整合，具体操作方法参照 Cochrane 干预措施评价手册。

(五) 质量控制

为了确保综述结果的可靠性，需要对操作环节进行质量控制，可参考如下操作：确认每一个操作过程是否正确规范并保存相关操作记录以备查验；对每一篇纳入文献进行质量评价；最终纳入的文献结果需由至少两名有相关经验的人士进行独立评价，对争议性的文献有必要进行审核复评，以使出现偏倚的概率降至最低并确保结果一致性；对未能达成一致的评价结果，则需另外具有相关经验的研究者在该研究被纳入证据基础前对其进行质量评判。

第四章 基于多准则决策分析的药品临床综合评审流程

以合理分配有限资源为宗旨，对不同的抗肿瘤药物比较分析时，需评估者在不同评价准则中进行取舍。多准则决策分析（Multi-Criteria Decision Analysis, MCDA）通过一系列方法对卫生技术（如抗肿瘤药物）的综合价值进行排序，帮助决策者在多种准则中作出取舍，从而确定最佳选择。

在确定维度和指标框架的基础上，评审专家以药品多维度价值判断操作流程为工具对待评估卫生技术给予相应的评审工作，主要分为如下环节（见图7）：

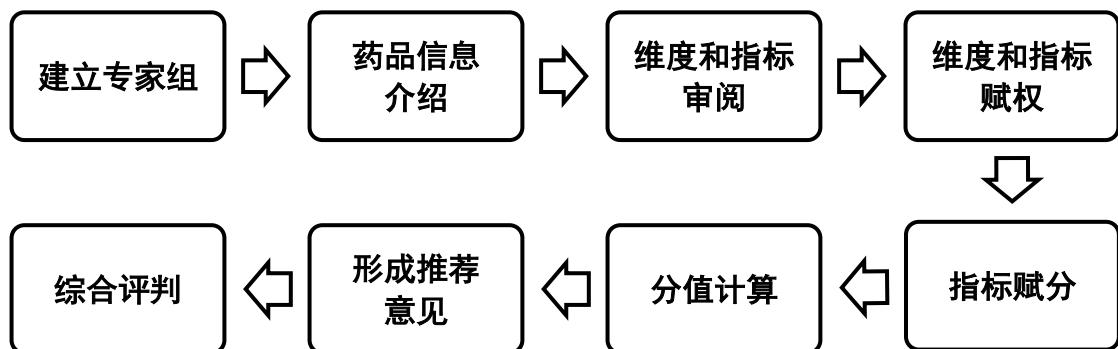


图7 药品多维度价值判断操作流程

一、建立专家组

首先根据目标药品从已建立的专家库中随机抽取一定数量适宜专家形成专家组，专家人数以奇数为主。专家入选标准为：①具备高级专业技术职称或在所从事的领域有一定知名度②本人或家属与所评审药品没有利益冲突。专家构成需突出多学科特点，主要包含卫生政策、医疗保险、卫生经济、药物经济、卫生统计、临床医务人员、临床药师、行政管理人员、患者。推举1名权威且富有经验的专家作为评审组长，进行后续环节的组织与监督，但不参与后期赋权评分工作，以备出现不同评审意见时行使最终决策权。

二、抗肿瘤药物信息介绍

在确定评审组长后，请抗肿瘤药物综合评价实施者就上述步骤整合的抗肿瘤药物多维度价值进行汇报，以便评审专家了解抗肿瘤药物具体信息，并接受与抗肿瘤药物相关问题的问询。

三、维度和指标审阅

鉴于评审为多学科、多利益相关者的融合的环节，评审组长需解释评审维度和评审指标内涵，评审专家可以讨论方

式对评审维度和指标做出研判，达成一致认知，为下一步权重评分提供基础。

四、维度与指标赋权

在确定评审维度和指标的基础上，评审专家依据自身专业判断和评估报告证据对维度与指标权重直接赋值。通过统计分析汇总整理专家意见，排除异常值²后，最终获得较一致、可靠的维度及指标的平均权重。

五、指标赋分

在确定维度和指标权重的基础上，每位专家需根据相对重要性，给每个指标进行指标赋分，借鉴李克特量表^[16]将重要性分为不重要、不太重要、一般、比较重要以及非常重要5个等级，赋值1-5分。评审专家根据药品临床综合评价报告相关抗肿瘤药物信息及主观判断，对指标进行赋分。例如：经济性维度下的增量成本效果比指标，可依据如下等级给出分值：□1 很差 □2 较差 □3 相同 □4 较好 □5 很好。通过对指标评分结果均值和异常值的判定，运用统计分析方法汇总整理专家意见，以获取每一指标对应分值。

² 异常值排除标准：下限：第一四分位数-1.5*四分位距；上限：第三四分位数+1.5*四分位距

六、分值计算

根据维度和指标权重及指标分值，去除异常值后的各指标评分的算术平均数乘以相应的指标权重和维度权重即得指标的加权分值。每一指标都将获得一个分值，将其相加即为评审药品最终评审得分，并形成推荐意见。

七、形成推荐意见

根据药品的作用机制、靶点、适应证、对照药品选择等不同，由评审专家组确定评判标准，依据最后的评审分值，可以形成不同强度的推荐意见，一般分为 A、B、C、D 四类：

①评审结果证据充分、结果确定的为 A 类，建议可直接按程序转化为基本临床用药管理相关政策结果；②评审结果证据比较充分、结果明确的为 B 类，建议按程序有条件转化为基本临床用药管理相关政策结果；③评审结果有一定证据支持、部分结果明确的为 C 类，建议在一定区域范围内或特定医疗机构内按程序转化为基本临床用药管理相关政策结果；④评审结果证据不足、结果不确定的为 D 类，不建议转化政策结果。

八、综合评判

药品临床综合评价实施小组应根据药品治疗疾病领域的特点、目前该疾病领域治疗药品在我国上市情况及医疗保障现状，基于多准则决策分析的药品临床综合评审内容，对药品进行有侧重点及针对性的综合评判。若待评价抗肿瘤药品在国际及我国上市较长，临床实践经验较为丰富，则可弱化其在创新性维度上的证据整合；反之，若待评价抗肿瘤药品治疗疾病目前存在临床空白，则该药品的安全性、有效性、经济性、创新性、适宜性、可及性等维度都应该详尽整合证据。抗肿瘤药品临床综合评价内容及维度的侧重点可根据药品及疾病特点进行调整，但需要描述调整的依据。

参考文献

- [1] CANCER I A F R O. China Cancer [Z]. 2020
- [2] 《中国药物经济学评价指南》课题组. 中国药物经济学评价指南(2019版) [M]. 2019.
- [3] KALTENTHALER E, COOPER K, PANDOR A, et al. The use of rapid review methods in health technology assessments: 3 case studies [J]. BMC Medical Research Methodology, 2016, 16.
- [4] A N L V D P E C. Practical measurement of affordability: an application to medicines [R]: Rotterdam: Institute of Health Policy and Management, Erasmus University, 2009.
- [5] 布拉德伯恩 萨, 万辛克. 问卷设计手册: 市场研究、民意调查、社会调查、健康调查指南 [M]. 重庆大学出版社, 2010.
- [6] 诺曼·布拉德伯恩, 希摩·萨德曼, 布莱恩·万辛克. 问卷设计手册: 市场研究, 民意调查, 社会调查, 健康调查指南: the definitive guide to questionnaire design for market research, political polls, and social and health questionnaires [M]. 问卷设计手册: 市场研究、民意调查、社会调查、健康调查指南: the definitive guide to questionnaire design for market research, political polls, and social and health questionnaires, 2011.
- [7] KHANGURA S, KONNYU K, CUSHMAN R, et al. Evidence summaries: the evolution of a rapid review approach [J]. Systematic Reviews, 2012, 1.
- [8] INAHTA. htaglossary [Z]. 2020
- [9] LYSDAHL K B, MOZYGEMBA K, BURNS J, et al. Comprehensive assessment of complex technologies: integrating various aspects in health technology assessment [J]. International Journal of Technology Assessment in Health Care, 2017: 1.
- [10] CADTH. About the Health Technology Assessment Service [Z].
- [11] 高培, 王杨, 罗剑锋, et al. 基于真实世界数据评价治疗结局研究的统计分析技术规范 [J]. 中国循证医学杂志, 2019, 019(007): 787-93.
- [12] International Clinical Trials Registry Platform [J/OL] 2015, <http://apps.who.int/trialsearch/>.
- [13] CADTH. Grey Matters: A Practical Tool for Searching Health-Related [J]. 2013.
- [14] HEDGES. 2016, https://hiru.mcmaster.ca/hiru/HIRU_Hedges_home.aspx.
- [15] ATKINS D, ECCLES M, FLOTTORP S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group [J]. BMC health services research, 2004, 4(1): 38.
- [16] 杨林, 陆阳. 基于层次分析法和德尔菲法的医疗设备采购评价方法 [J]. 中国医疗设备, 2017, v.32(12): 185-8.

中英文对照表

英文缩写	英文全称	中文全称
CBA	Cost-Benefit Analysis	成本-效益分析
CEA	Cost-Effectiveness Analysis	成本-效果分析
CR	Complete Response	完全缓解
CRF	Case Report Form	病例报告表
CTCAE	Common Terminology Criteria for Adverse Events	常见不良反应事件评价标准
CUA	Cost-Utility Analysis	成本-效用分析
DDFS	Distance Metastasis free survival	无远处转移生存期
DFS	Disease-free Survival	无病生存期
DID	Difference in Difference	倍差法
EORTC QLQ	European Organization for Research and Treatment of Cancer Quality of Life Core Questionnaire	欧洲癌症治疗研究组的肿瘤生存质量调查问卷
EQ-5D	EuroQoL Five Dimensions Questionnaire	欧洲五维健康量表
HAI	Health Action International	国际健康行动机构
HIS	Hospital Information System	医院信息系统
ICER	Incremental Cost Effective Ratio	增量成本-效果比
ISPOR	International Society for Health Economics and Outcomes Research	国际药物经济学会
ITS	Interrupted Time Series	间断时间序列分析
MPR	Median Price Ratio	中位价格比
NCCN	National Comprehensive Cancer Network	美国国立综合癌症网络
ORR	Objective Response Rate	客观缓解率
OS	Overall Survival	总生存期

英文缩写	英文全称	中文全称
PCT	Pragmatic Clinical Trial	实效性临床试验
PD	Progressive Disease	疾病进展
PFS	Progression-free survival	无进展生存期
PR	Partial Response	部分缓解
PRO	Patient-reported Outcome	患者报告结局
PSM	Propensity Score Matching	倾向评分匹配
PSM	Partitioned Survival Model	分区生存模型
QALYs	Quality-Adjusted Life Years	质量调整生命年
RCT	Randomized Controlled Trial	随机对照临床试验
RECIST	Response Evaluation Criteria in Solid Tumors	实体瘤疗效评价标准
SD	Stable Disease	疾病稳定
TTP	Time to Progression	至疾病进展时间
WHO	World Health Organization	世界卫生组织

《抗肿瘤药品临床综合评价技术指南》编委会

主编单位：国家药物和卫生技术综合评估中心（国家卫生健康委卫生发展研究中心）

责任编辑单位：国家癌症中心，国家卫生健康委药具管理中心

指导单位：国家卫生健康委药物政策与基本药物制度司

抗肿瘤药品临床综合评价技术指南编写组（单位不分先后）：

国家药物和卫生技术综合评估中心、中国医学科学院肿瘤医院（国家癌症中心）、哈尔滨医科大学附属肿瘤医院、北京大学人民医院、山东省立医院、北京大学肿瘤医院、首都医科大学、天津大学、中国疾病预防控制中心、首都医科大学附属北京友谊医院、复旦大学附属肿瘤医院、北京大学第三医院

附件 标准报告格式

项目	内容要求
一、报告首页	
1、标题	标题为“XXX 治疗 XX 肿瘤疾病临床综合评价报告”，其中 XXX 为被评价药品通用名
2、评价单位	完成综合评价的单位
3、评价报告完成日期	完成综合评价的日期
二、摘要	
1、目的	拟采取何种设计通过评价哪些维度，拟解决的临床用药监管管理决策问题
2、评价方法	说明评价设计及主要数据收集方法及工具
3、评价结果	回答评价目的所针对的临床用药监管管理决策问题
4、评价结论及建议	经专家论证后的评价结果结论及建议
三、正文	
1、评价背景	简述评价背景，包括所针对疾病流行病特点及用药现况，介绍评价主题提出的意义。
2、评价目的	描述通过什么评价方法，拟评价药品应用的哪些问题，旨在解决何种用药监管管理问题
3、评价过程	具体描述评价流程及评价方法，描述应按照实际过程和步骤，每个步骤具体如何进行，得到何种具体结果，使用与指南不同的方法时应注明原因。按照评价流程及评价方法具体过程，列表写明评价步骤及相应的评价人员，评价起止日期，评价原始记录与数据保存位置，是否质控，质控人员，质控起止日期，专家评审及结果。

项目	内容要求
4、评价结果	评价药品与对照之间的数据统计分析结果。此处应列出所有方法的分析结果，不同方法分析结果可能不同，结果相悖处应使用*号注明，并在后面加括号，括号中注明与何处结果相悖
(1) 技术特性	主要有效成分： 国外注册时间： 国内注册时间： 进口企业情况： 同族中国专利及到期日： 国产制剂申报企业数： 国产制剂获批企业数： 与原研药比较（有效成分，药代药动，稳定性，给药方式）： 与同类仿制药比较（有效成分，药代药动，稳定性，给药方式）： 最优可替代药品为：
(2) 健康需求	适应症： 疾病负担： 目前治疗模式： 临床需求：
(3) 政策重要性	是否为临床基本用药（基本药物目录，基本医保目录用药）： 是否临床急需用药： 是否存在供应短缺： 是否为公共卫生应急用药： 是否是国家卫生健康创新产品：
(4) 安全性	与对照药品比较安全性及相对安全性
(5) 有效性	与对照药品比较有效性及相对有效性

项目	内容要求
(6) 经济性	评估药品价格及治疗方案费用 与对照药品比较成本效果及相对成本效果
(7) 适宜性	包括疗程长度、药品配伍、应用时间（间隔）、 用药监测、临床效果、经济性等相关适宜性
(8) 创新性	是否具有临床创新性、服务体系创新性、产业创 新性
(9) 可及性	价格、可获得性、可负担性
(10) 其他证据	如果还调取其他证据，请在此处汇报
5、评价结论	评价专家指导组给出的详细建议和评价结论，注 明详细理由。对于评价结果相悖的，应写明评价 结果不同的原因，采信某结果的理由，并给出未 采信结果的理由及证据。
四、参考文献	
五、主要附件 (1) 主要技术附件（包括二手数据来源、检索策略、检索结果和纳入研究 列表；调查问卷；访谈提纲；专家咨询记录；会议纪要等） (2) 主要管理附件（包括团队名单，内外部质控评价结果）	