

AASLD-EASL 联合研讨会关于 NAFLD 临床试验研究终点的报告

李静¹ 译, 吴瑞红² 牛俊奇² 审校

1 广东众生睿创生物科技有限公司; 2 吉林大学第一医院

2018年6月美国肝病研究协会(AASLD)与欧洲肝病研究协会(EASL)召开了非酒精性脂肪肝病(NAFLD)临床试验研究终点的联合研讨会。会议总结了已完成及进行中临床试验的重要发现,确定了支持不同临床试验终点的科学证据,为今后的试验设计提供指导。本文主要报告会议的基本意见及共识观点。

摘要

非酒精性脂肪肝病(NAFLD)是一个全球性的公共健康问题。其自然史,非酒精性脂肪性肝炎(NASH)和纤维化的形成呈现出高度异质特点,易受内源性因素(如遗传学、微生物群)和外源性因素(如营养、酒精、体育活动)的影响,并可随时间的变化而波动。NAFLD病理生理学的复杂性也体现在纷繁复杂的NASH研发药理靶点上。已有的NASH临床试验为今后的临床设计提供了有价值的经验。NASH的研究终点在过去十年中不断发展,并将持续改进。目前可接受的有条件批准的终点包括NASH缓解且纤维化无恶化和/或纤维化改善且NASH无恶化,需通过治疗前后配对的标准化的肝脏组织学评估。但在儿童NASH中,由于操作问题、青春期激素变化和严格的安全性要求,在试验设计中需进行适当调整。在NASH相关成人肝硬化患者中,临床事件(如失代偿、肝细胞癌、移植、死亡)较为常见,因此可作为主要终点。考虑到疾病的自然波动,所选主要终点的临床意义,以及可能影响安慰剂反应的因素,将有助于准确判断新疗法对NASH的疗效。

前言

非酒精性脂肪肝病的疾病谱包括从非酒精性脂肪肝(NAFL)到NASH伴进展性纤维化的一系列疾病^[1-5]。肝纤维化与NASH紧密相关,可进展为肝硬化及终末期肝病^[4]。表1概述了目前被认可的NASH组织学分期分级评价系统。NASH与代谢综合征的一些临床特征和严重程度之间也存在共线性(collinearity)关系^[6-8]。NAFLD是心血管疾病发生的独立危险因素^[9]。

目前尚无被批准用于治疗NASH的药物^[10]。在美国,药物的批准依据确证的临床获益或可被食品药品监督管理局(FDA)接受的替代终点的改善。欧洲药品管理局(EMA)的标准与之类似。而确证临床获益可能需要数年时间,因此美国和欧洲都建立了有条件批准体系,视情况采用“合理可接受(reasonably accepted)”的替代终点。使用这些替代终点的一个关键挑战是它们的定义和支持其有效性和安全性证据的生成。AASLD与EASL于2018年6月29日至30日在维吉尼亚州亚历山德拉市(Alexandra, Virginia)召开了关于NAFLD临床试验终点的联合会议。会议对NASH临床试验的科学依据,相关知识差距和试验终点进行了综述,主要目的是指导和改进NASH临床试验设计和实施。本文主要报告会议的重要内容及共识。

表 1 临床试验中当前被认可的 NASH 分期和分级评分系统

(Currently Accepted Scoring Systems to Grade and Stage NASH in Clinical Trials)

欧洲脂肪肝进展阻断组织学评分 (FLIP Steatosis, Activity, and Fibrosis Score)				美国糖尿病、消化和肾脏疾病研究所 脂肪变性、活动性和纤维化评分 (National Institute of Diabetes and Digestive and Kidney Diseases NAS)		
组织学特征	分值范围	分级	分级定义	分值范围	分级	分级定义
脂肪变性	0-3	0	<5%	0-3	0	<5%
		1	5%-33%		1	5%-33%
		2	34%-66%		2	34%-66%
		3	>66%		3	>66%
SAF 脂肪变评分 (S) : 0-3						
肝细胞气球样变	0-2	0	无	0-2	0	无
		1	圆形胞浆透明的肝细胞团		1	少量
		2	在 1 分基础上肝细胞增大 (>2 倍正常大小)		2	很多
炎症	0-2	0	无	0-3	0	无
		1	≤2 个病灶/20 倍视野 *		1	<2 个病灶/20 倍视野 #
		2	>2 个病灶/20 倍视野		2	2-4 个病灶/20 倍视野
					3	>4 个病灶/20 倍视野
SAF 活动度评分 (A) : 气球样变+炎症: 0-4				NAS 总分=脂肪变+气球样变+炎症: 0-8		
纤维化		0	无纤维化		0	无纤维化
		1a	3 区窦周轻度纤维化		1a	3 区窦周轻度纤维化
		1b	3 区窦周中度纤维化		1b	3 区窦周中度纤维化
		1c	汇管区纤维化		1c	汇管区纤维化
		2	3 区+汇管区		2	3 区+汇管区
		3	桥接		3	桥接
		4	肝硬化		4	肝硬化
	纤维化评分: 0-4			纤维化评分: 0-4		

FLIP: 脂肪肝进展阻断评分系统; SAF: 脂肪变性, 活动性和纤维化评分; NAS: NAFLD 活动性评分

* 译者注: 此处翻译基于的参考文献为 Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. Bedossa P, Poitou C, Veyrie N, et al. Hepatology. 2012 Nov;56(5):1751-9.

译者注: 此处翻译基于的参考文献为 Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Kleiner DE, Brunt EM, Van Natta M, et al. Hepatology. 2005 Jun;41(6):1313-21.

流行病学

从 2005 年到 2010 年，NAFLD 全球患病率从 15% 增加到 25%，预计到 2030 年将进一步增加，并且进展期患者所占比例将更高^[11-12]。NAFLD 的患病率及严重程度因地区和种族而异，患病率最高的在中东（31.79%）、南美（30.45%）和亚洲（27.37%），其次是北美（24.13%）和欧洲（23.71%），非洲最低（13.48%）^[2]。

NAFLD 在不同地理区域和种族间的患病率和严重程度的差异有许多可能的原因。NAFLD 明显受遗传因素影响，但它是一种复杂的疾病（complex disease），疾病严重程度还受环境及行为因素影响（如饮食、体育活动、社会经济）。影响疾病严重程度最具特征的遗传因子是编码 Patatin 样磷脂酶域蛋白 3（PNPLA3）基因上的 rs738409 c.444 C>G（p.I148M）和编码跨膜蛋白 6 超家族成员 2（TM6SF2）基因上的 rs58542926 c.449 C>T（p.E167K）。它们与脂肪变程度、NASH 及纤维化/肝硬化密切相关^[13,14]。最近研究表明膜结合 O 型酰基转移酶 1（MBOAT1）基因上的 rs641738 以及 17-β 羟基类固醇脱氢酶 13（HSD17B13）基因上的 rs72613567 也与 NAFLD 有关^[15,16]。

这些遗传多态性在各种族间的分布存在很大的差异。各大洲间潜在异质性对 NASH 的临床试验设计和药物开发具有重要意义，尽管基因对安慰剂反应和药物治疗应答的影响尚不清楚。在关键性试验中，通过区域分层或不同地区进行单独试验，可能消除地理上遗传多样性效应。

NAFLD 的自然史和亚型

NAFLD 的不同的疾病状态之间存在动态变化，NASH 甚至纤维化都可以随着时间的推移而进展、逆转或保持稳定。这些结果可能会受到生活方式的强烈影响，如营养、体育活动和饮酒^[17,18]。虽然 NAFLD 的患病率很高，但仅小部分患者发展成为进展性肝病或引起肝脏相关死亡。NAFL 向 NASH 的转变具有很强的动态性和双向性，而且这种来回转变的频率和因此产生的损害程度尚未得到充分描述^[3,4]。此外，纤维化进展的中位速率以及快进展和慢进展亚群具有的特征也尚未充分确定（图 1）。活检取样误差，病理学专家之间/甚至同一专家在阅片上的差异，以及纤维化的明显自发改善等情况变得更加复杂。因此，已有报道指出应用安慰剂的患者中获得组织学纤维化改善 1 个等级的约占 20%（15%-35%）^[4]。在一项又接着进行了后续活检的研究中，最初那些应用安慰剂 1 年后纤维化获得改善的患者，很高比例（80%）在第 1 年至第 2 年期间恶化^[19]。进展期纤维化（F3-F4 期）患者进展为不良结局（hard outcomes，肝脏疾病失代偿，肝移植，死亡）的风险急剧增加。因此，很明显优先考虑预防进展为进展期肝纤维化或诱导其逆转。进展期肝纤维化，特别是确诊肝硬化，相比早期纤维化不那么容易发生自发性逆转。对于疾病晚期阶段，药物诱导纤维化逆转可能是一个特别艰难的终点^[17,18]。

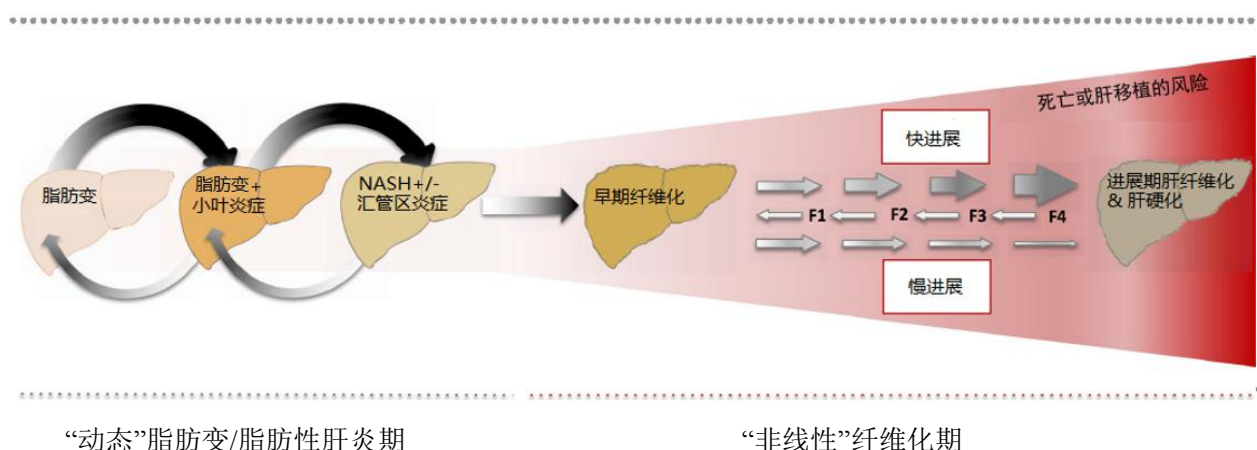


图 1: NAFLD 自然史动态变化。基于纵向队列研究和两次活检队列的观察，NAFLD 的自然史是高度动态变化的。脂肪性肝炎的严重程度即 NAFL 和 NASH 之间的转化具有周期消长的特征。同样，纤维化也可能进展或逆转，虽然大多数患者进展缓慢，但多达 20% 可能是快进展患者。发病率和死亡率随纤维化程度的增加而增加。

并存病 (comorbidities)

NAFLD 患者通常伴随肝外疾病，如肥胖、胰岛素抵抗、2 型糖尿病、代谢综合征和心血管疾病等。并存病的存在可能影响治疗的效果。例如，NASH 患者合并糖尿病可能对吡格列酮产生更强的反应，而安慰剂反应低一些^[20]。

因此，在临床试验设计时，并存病如 2 型糖尿病等可能值得考虑，可视药物的作用机制而定。目前，尚无足够的证据将代谢性并存病的某些方面纳入试验终点。由于 NASH 患者心血管事件的几率增加，目前明确的共识是治疗 NASH 的药物不应增加心血管疾病的风险。一般来说，治疗 NASH 的药物应对代谢性并存病无影响，或最好能产生有益的效果。有限的数据提示 NAFLD 活动性评分 (NAS) 的降低与心血管风险的改善有关^[21]。建议仔细评估治疗药物对心脏-代谢危险因素的影响，将其作为 II 期和 III 期试验的次要终点。

酒精作为潜在辅助因子在 NAFLD 进展中的作用

当前临床试验设计中没有充分考虑的一个潜在的重要因素是酒精量、饮酒持续时间和饮酒模式对疾病进展或自发改善的影响。临床试验排除每天饮酒超过 20/30g (女性/男性) 的受试者；然而，这个数值是根据高于此水平时肝硬化风险会更高界定的，没有明确表明其是否适合于 NASH^[22]。此外，酒精与 NAFLD 等其他肝病同时存在时，会加重肝损伤^[17]。

目前，酒精摄入量的评估通常仅限于试验开始时的筛查，既没有应用生物标志物 (如尿乙基葡萄糖醛酸苷、磷脂酰乙醇)，也没有用临床实践或试验中常规使用的饮酒危险的标准心理测试。更准确地评估酒精摄入量的必要性被低估了；因此，更好的评估可以识别和控制这一重要混杂因素。

对基线组织学严重程度相同的个体进行前瞻性随访发现，随着时间的推移，那些适度饮酒的人脂肪变性程度更高，或改善程度更低，而在炎症、气球样变或纤维化方面没有差异^[23]。这并不排除少量饮酒可能会改变治疗结果的可能性。酒精时间表跟踪评估 (Alcohol Timeline Follow-back assessment) 是目前跟踪饮酒量的“最佳实践”，但需要对人员进行特定的培训^[24]。可以考虑在长期试验的背景下评估其如何影响试验结果。

安慰剂反应 (Placebo Response)：试验成功的一个关键决定因素

最终，研究化合物的疗效是否优于安慰剂决定了干预的成功与否。丙型肝炎病毒感染的临床试验中安慰剂反应可以忽略不计，而 NASH 试验中的安慰剂反应在不同的研究中显著不同。在最近报道的 NASH 研究中，安慰剂组织学反应范围在 9%-40% 之间^[25-29]。最近的一项荟萃分析进一步量化了这种效应，39 项随机对照试验，纳入 1463 名安慰剂组患者^[30]，25% (95% 置信区间 [CI] 21%-29%) 患者出现 NAS 评分下降 2 分，超过 1/3 的患者在至少一个 NAS 评分组分中 (脂肪变性、气球样变或炎症) 改善超过 1 分。此外，21% (95% CI 16%-26%) 患者表现为纤维化改善至少 1 级^[30]。

影响安慰剂反应的因素可以大致分为“患者相关”和“研究设计相关”。一个经常被低估的与患者相关的因素是“霍桑效应 (Hawthorne effect)”，即一个人意识到被观察，或者只是参与了临床试验，就会改变他的行为^[30-32]。这种现象与 NAFLD 等疾病尤其相关，生活方式的改变可以显著影响基础疾病。

生活方式对 NAFLD 临床试验结果的显著影响，导致很难确定试验结果是由于药物治疗还是生活方式的改变引起。生活方式的改变甚至产生一定的效果 (例如，持续减重 $\geq 5\%$ 可以降低脂肪变性^[7]、肝酶^[31]和代谢指标，有时被用作 NAFLD 临床试验的次要终点)。此外，除了体重减轻，膳食结构本身 (地中海式 vs. 西式) 也是影响肝脏预后的一个因素^[19]。纤维化对体重或膳食结构的轻微变化不太敏感，但在体重持续下降超过 10% 的患者中出现逆转^[31]。因此，建议对饮食 (热量摄入、成分以及如前所述的酒精摄入量)、体育活动和人体测量参数 (体重、体重指数、腰围、内脏脂肪组织、腹部脂肪) 进行数据采集。对这些数据进行分析可以发现潜在的混杂效应、组间不平衡，并有助于在试验中进行比较^[32]。

可考虑采用多种技术，包括协助自我监测体力活动和/或热量摄入的移动应用程序，以便更准确地量化参与临床试验的个体采纳生活方式建议的程度^[33]。数据跟踪和获取有可能阐明哪些因素影响疾病进展，并确定检测到的健康指标改善是否可归因于生活方式因素还是药物治疗^[32]。还应考虑改变研究设计，在进行基线活检和开始积极干预之前纳入导入期，以减轻生活方式改变的影响和/或确定两组是否受到相似的影响。

研究设计也可能会显著影响安慰剂反应。这主要与试验入组标准有关，特别是入组的组织学严重程度阈值和采用的疗效终点的严格程度。大多数研究都明确了临床试验的最低 NASH 分级和纤维化分期。过去的试验提供了宝贵的经验教训，证明宽松的纳入标准可以影响试验结果，主要表现在安慰剂反应的增高。在 GOLDEN-505 实验中，NAS \geq 4 的受试者的安慰剂反应率（PRR）低于 NAS 评分较低的受试者（基线 NAS=3 的患者 PRR 为 25%，NAS \geq 4 的 PRR 为 9%）^[26]。因此，采用更严格的 NASH 缓解终点的定义可以降低安慰剂的反应率^[5]。

对于纤维化相关终点来说，以上观点虽然在文献中定义的不太明确，但也可能成立，在于安慰剂组纤维化自发改善达到 1 个等级（第一年 10%，第二年 17%）比改善达到 2 个等级（第一年 3%，第二年 3%）更有可能发生^[26]。

识别 NAFLD 潜在靶向途径的临床前工具

广泛的研究工作已经确定了多条导致疾病的通路，这些通路可作为创新药或“老药新用”的靶点。进入临床开发所需的临床前证据的水平备受争议。为此，开发了若干创新的体外模型（如肝切片、生物芯片、类器官）以及在活体（in vivo）啮齿动物模型（如遗传、化学、饮食）。关键的挑战是目前没有任何一种模型可以全面模拟人类疾病的所有方面。然而，人类 NAFLD 的不同特征可以在小鼠模型中复制，包括脂肪变性、炎症和纤维化。人们一致认为，没有一种完美的模型能够给所有作用机制的干预措施的有效性提供最优评价。最佳方法是，合理选择最能反映新化合物靶向致病因素的模型。已发表的 II 期试验文献表明，临床前的有效性无法准确地反映到患病人体上。其原因很大程度上可能是由于人类 NAFLD 的复杂性，这种复杂性是指在特定个体中，受外部因素影响的主导机制存在异质性。另一方面，疗效的减弱或缺乏可能与临床前数据不足以支持进入 II 期临床试验有关。我们的首要责任是保证患者的安全与健康，我们必须获得尽可能有力的临床前数据来支持进入人体临床试验。以下原则可能有助于预测临床前模型中出现信号的临床相关性：

1. 疗效信号应在一个以上实验模型中保持一致。理想情况下，疗效应在不同的实验室和条件下（如不同的微生物群、小鼠品系、处理方法）重现；
2. 实验模型应正确反映药物靶向的致病机制，这种具体作用机制应与人类疾病相关；
3. 药物干预也应在“恢复模型”中进行测试（如将小鼠从致病性饮食改为正常饮食以诱导疾病恢复），而不延迟疾病逆转/缓解；
4. 对疾病状态下小鼠和人体进行干预的时机应该相似，药物在组织器官的暴露剂量和靶向作用应该可比，临床前研究应设计合适的随机方法和能检测疗效差异的把握度。

临床试验及数据采集的建议标准

现有的 III 期临床试验将为根据自然史和对治疗药物的应答对 NASH 患者进行分层提供重要的初步证据。对 NASH 患者进一步分类将改进未来的研究设计，并有助于确定不同治疗方案的最适合患者。迫切需要更多的表型分析工具和确证的生物标志物，来进一步细分临床试验中的 NASH 人群。这些挑战正在通过大规模的国际学术界和工业界合作来解决，例如美国国立卫生研究院的代谢性肝病非侵入性生物标志物基金联合会（Noninvasive Biomarkers of Metabolic Liver Disease consortium of the Foundation for the National Institutes of Health in the United States）^[34]和欧洲脂肪性肝炎标记物检测联盟的肝脏调查

研究项目（the Liver Investigation: Testing Marker Utility in Steatohepatitis consortium in Europe）^[35]。遗传标记物，肠道微生物群的变化（生态失调），代谢（脂质标记物、胆汁酸、肠道代谢物）、胶原代谢以及影像学技术可能有助于脂肪肝患者表型的分类。因此，临床试验提供了一个收集证据的机会，以支持针对 NASH 患者的个性化医疗方法。在 III 期临床试验中，可以获得种族或基因差异数据，可以使用单核苷酸多态性芯片或测序技术来寻找易于或不易于对治疗有反应或有不同药物副作用风险的疾病亚型^[5]。然而，治疗性试验设计的初衷并非评估这些因素的作用，因此，需要在进一步的研究中独立验证相关性。

NAFLD 临床试验的结果报告

II 期临床试验结果报告的不一致性限制了不同研究的结果比较和对新疗法的安全性和有效性的客观评价。遵守 NAFLD 临床试验结果报告的标准规则将提供帮助有助于了解化合物在该人群中的疗效。一个强烈的共识是，NAFLD 临床试验的结果报告应该遵循国际医学期刊编辑委员会（International Committee of Medical Journal Editors）的规则，具体而言应该：

1. 包括患者基线特征的详细描述，不仅包括人口统计学和肝病状况，还包括共存病（如肥胖、代谢综合征、2 型糖尿病、心血管疾病、抑郁症），饮酒量的定量评估以及伴随用药^[22]；
2. 报告该研究的所有主要终点和次要终点相关的疗效数据，报告所有疗效相关结果的绝对变化（包括 *P* 值），而不仅仅报告变化的百分比；
3. 报告接受治疗药物和安慰剂后病情改善、稳定或恶化的绝对人数和比例；
4. 报告与研究终点相关的各诊断方法和生物标志物之间的一致性或不一致性（例如，同时报告纤维化组织学分期和无创纤维化评估结果）；
5. 避免使用“有临床意义的改变”之类的用语，除非有研究队列中报告的临床结果数据支持；
6. 报告详细的安全性数据（不良事件、药物引起的肝损伤、肝脏相关事件、肝移植、心血管事件、死亡），特别是包括对胰岛素抵抗/血糖控制和循环血脂的影响。

研究终点

目前公认的对于肝硬化前（precirrhotic）NASH 有条件批准的终点包括 NASH 缓解同时纤维化无恶化和/或纤维化改善同时 NASH 无恶化^[36,37]。表 2 根据 NASH 谱的发展阶段概述了当前可接受的终点。诸如“NASH 改善”（以及如何定义它）等变量的效用仍需要进一步研究。要注意的是终点的设定必须结合给定药物的作用机制来考虑。对于作用机制主要为代谢或抗炎的药物，预期脂肪变性和肝细胞损伤（气球样变和炎症）改善。然而，对于一种纯抗纤维化药物，虽然预期可能会产生不依赖于脂肪变性或疾病活动度改善的抗纤维化作用，但不应对疾病活动性产生负面影响。

目前，纤维化改善达 1 个等级是可接受的终点。然而，随着更多证据的积累，这可能需要修正，因为在 3 个最大的 II 期研究中纤维化改善达 1 个等级不尽相同，并且安慰剂反应从 14%到 44%不等。采用更严格的标准，如纤维化改善至少 2 个等级，将增加特异性，并将安慰剂反应的不可预测性降到最低。然而，如果按此标准实施，达到主要终点的受试者比例可能减少，这可能会增加样本量要求和试验持续时间。

关键临床结局（Key Clinical Outcomes）

NASH 合并进展期肝纤维化仅占 NAFLD 患者少数，肝脏相关性死亡为其最显著的死因。迄今为止最大的长期队列研究发现肝硬化前 NAFLD 患者死亡的主要原因为心血管疾病（38%）和肝外恶性肿瘤（19%），肝脏相关死亡（含死于肝细胞癌 HCC）占总病例的 9%^[39]。最近一项工作对更细组别的患者，进展期肝纤维化（F3）和代偿期肝硬化（Child-Pugh A5 分或 A6 分）NAFLD 患者，报告了其临床结局^[40]。不足为奇，这项研究显示心血管疾病的年发病率在肝硬化前 NAFLD 患者较高，在肝硬化 Child-Pugh

A6 患者最低(0.9/年[95% CI 0.5~1.8] VS. 0.2 /年[95% CI 0.03~0.6])^[40]。而全因死亡/肝移植年发生率(0.5 [0.2-1.2] vs. 11.1 [8.3-14.8])及肝癌的年发生率(0.2 [0.02-0.9] vs. 4.7 [3.0-7.5])在这两组呈现相反趋势^[40,41]。虽然心血管疾病和 NASH 之间存在联系, 而且有证据表明 NASH 的缓解与促动脉粥样硬化的血脂有关, 但目前没有数据表明逆转 NASH 可以影响心血管结局^[21]。肝脏相关的结局包括死亡、移植(或用终末期肝病模型[MELD]评分>15 替代)、失代偿期肝硬化(腹水、静脉曲张出血、肝性脑病)和 HCC, 也更倾向发生在进展期肝纤维化患者中, 尽管 HCC 可以发生在 NAFLD 的任何阶段^[1,4,8-10]。

对代偿期肝硬化患者进行分层, 以鉴别出可能发展为失代偿期肝硬化的亚群仍存在相当大的挑战。需特别提下, 门脉高压^[12](肝静脉压力梯度[HVPG]≥6 mmHg), 尤其是 HVPG>10 mmHg, 是肝硬化失代偿的一个关键预测因素^[41]。HVPG>10 mmHg 的患者 HVPG 每增加 1 mmHg, 肝失代偿风险增加 11%。HVPG<10 mmHg 意味着在中位随访 4 年期间有 90%的可能性不发生临床肝功能失代偿^[42]。但这个阈值在仅限于 NASH 人群时, 可能是不同的。新生临床试验数据可以让我们对它进一步认识。首次治疗 NASH 肝硬化患者的临床试验结果显示, 这个阈值在 NASH 中可能没那么可靠, 因为一些 HVPG<10 mmHg 的患者出现了门脉高压并发症。一旦发生失代偿, MELD 评分可能是一个比 HVPG 更优的死亡预测工具^[43]。MELD 评分>10 也能预测失代偿风险, 可能被用作分层肝硬化的工具; 然而, 新产生的数据显示这些截断值(cutoffs)在 NASH 肝硬化中可能不太可靠^[44]。此外, 在这一人群中, 各失代偿事件是否有相似的临床意义也尚不明确。一个正逐渐达成的共识是发生腹水对预后的影响大于仅发生静脉曲张出血^[44]。以往代偿期肝硬化患者发生临床失代偿的生物学机制尚不完全清楚。因此, 治疗潜在的脂肪性肝炎而不是纤维化是否能预防临床失代偿仍有待确定。目前正在积极研究这些可能性。此外, 一旦肝硬化发生, 一些生物学过程也会参与其中, 目前还不清楚 NASH 疗法是否会对这些过程产生积极的影响, 尤其是在肝失代偿情况下。预计需要对肝硬化患者进行进一步细分, 并明确终点, 以促进该人群的治疗发展。

对于肝硬化前的 NASH, 法规中的替代终点包括 NASH 缓解(resolution), 疾病活动的减少(尚未明确定义), 或纤维化分期的下降。此外, 还有一种“可能合适(likely)”的替代终点, 如进展到肝硬化(或从肝硬化逆转)等, 是被“普遍接受”用以代表那些有临床意义结局的替代指标(表 2)。肝外临床结局包括心血管疾病和肝外恶性肿瘤; 然而, 在当前的 NASH 临床试验背景下, 这些结局作为替代终点不太可行^[9]。

表 2 NASH 临床试验中推荐的终点指标

	肝硬化前 NASH (Precirrhotic NASH)	NASH 肝硬化	儿童 NASH
早期研究	ALT、AST 肝脏脂肪 (MRI-PDFF) 注：应考虑药物作用机制	安全性 肝损伤研究 机制验证研究；基于作用机制选择终点指标	无明确推荐
2b 期	NASH 缓解 (resolution) 而无纤维化进展，或纤维化至少降低 1 个等级 NAFLD NAS 降低至少 2 分： 小叶炎症或气球样变至少降低 1 分 并且 纤维化分期无进展	MELD 评分从 <12 增加至 ≥15 NASH 无恶化且纤维化改善至少 1 分 HVPg 下降 ≥2mmHg	无明确推荐
3 期	NASH 逆转 (resolution)* 且纤维化无进展和/或 NASH 无恶化而纤维化改善至少 1 分	MELD 评分从 <12 进展至 >15 纤维化减低至少 1 分且 NASH* 无恶化 肝脏失代偿 ● 腹水 ● 静脉曲张出血 ● 肝性脑病 败血症 全因死亡 肝移植† 住院率	无明确推荐
4 期	临床获益确证试验 综合： ● 组织学进展为肝硬化 ● 全因死亡 ● 肝移植† ● 肝功能失代偿 ● MELD 评分从 <12 增加至 ≥15		无明确推荐

注：将未成年人纳入临床试验，需要证明药物干预对受试者的潜在直接获益大于最小风险，药物干预可从成人研究、临床前研究和某些机制研究中获得。有足够的关于成人用药剂量、安全性和有效性的信息，完成风险/效益评估后，方可用于儿童患者研究。

*气球样变消失，小叶炎症消失或少量存在，不符合 NASH 的诊断。

†人们认识到肝脏移植依赖于大量与肝脏疾病无关的因素，如经济状况、行为问题、社会支持、依从性、合并症情况和严重程度，以及一个人生活的地区。学术界的立场是，应以满足考虑肝移植的肝病严重程度标准而不是以肝移植的数量作为考虑的标准。

‡在这一点上，作者与 FDA 存在分歧。在 FDA 最近的指南中，并不鼓励将逆转肝硬化作为这个时期的终点，尽管这可以个例进行讨论。ALT：丙氨酸氨基转移酶；AST：天冬氨酸转氨酶；MRI-PDFF：磁共振成像-质子密度脂肪分数

疾病恶化的定义

会议讨论了“NASH 恶化 (Worsening of NASH)”的概念；但尚未就其精准定义达成共识。在短期研究中^[1,4,10,26,39,40]，尚未发现单纯脂肪变性增加与纤维化的恶化相关，认为其也不能反映脂肪性肝炎的恶化。最近研究表明气球样变和炎症的加重，尤其是汇管区炎症，很可能与纤维化加重相关。随时间推移，综合疾病活动评分的增加也与纤维化恶化有关。必须指出，这些分析基于回顾性数据，其方法学是受争议的，而关键性临床试验 (pivotal trials) 的数据让我们更有把握。

NASH 在组织学上的变化是动态的，如何定义疾病的来回波动(waxing and waning of disease)不重要，重要的是对疾病的恶化达成共识。从法规的角度来看，纤维化分期增加 1 级代表纤维化恶化。然而，人们认识到，目前的纤维化分期没有准确地描述纤维化的连续性，原因在于分期的序数性质给处在分级界限边缘的评价带来了误差。此外，纤维化各分期的变动范围不是线性的，因而从 F0 到 F1 期和从 F3 到 F4 期虽然均表现为进展了 1 个等级，但是纤维化变动范围是不同的。纤维化分期会同时捕捉纤维化的量和纤维化的分布信息，这会产生内部分歧。因此，一位纤维化 F2 患者的细胞外基质沉积可能比一位 F1 患者少。纤维化量也不是线性增加的，人们注意到从纤维化 F2 到 F3 阶段，纤维化量有一个显著的上升。这对于测定纤维化量而非分布的生物标记物的验证有重要意义。值得注意的是，一些试验已经引入了自动量化活检组织 (*biopsies*) 中胶原沉积染色进行量化分析，但还没能将这个参数与病理学专家评价的纤维化分期相关联。

疾病稳定 (Disease Stability)

在临床试验中应报告阻断疾病进展 (疾病维持稳定) 的结果，特别是对于未达到进展期 (less-advanced) 的疾病阶段及有疾病进展风险的患者。关于纤维化，目前其研究终点是关注纤维化逆转，而与其相关的临床结局是进展为肝硬化。因此，疾病稳定 (即未进展为肝硬化) 也可以考虑作为一个终点，但尚未被监管机构接受^[4]。

肝硬化临床试验的终点

肝硬化患者可以有不同的终点，并且最可能出现不良结局 (**hard outcomes**, 如失代偿、进展到 MELD \geq 15、肝移植、死亡)。组织学纤维化分期和 HVPG 的改善都已经成为 NASH 肝硬化临床试验的主要终点。每种都有局限性，并且达到这些终点对临床结局的影响必须得到证实。例如，纤维化分期的改善可由真实的逆转或抽样偏差造成，而 HVPG 的变化可由真实的肝门静脉压力下降、操作误差或其他混杂因素造成。

尽管 HVPG 的变化可能与技术操作不一致有关，但在 IIb 期试验中，不太可能出现这种情况，因为在 IIb 期试验中，每个 HVPG 的测量都经过了质量检查。HVPG 既受肝纤维化变化的影响，也受肝窦阻力动态变化的影响。这些数据采集不够细，解释不了安慰剂组 HVPG 改善的原因。此外，虽然 HVPG 恶化的意义已很明确，但 HVPG 改善的意义被关联到了 HVPG 降低的机制 (例如，经颈静脉肝内门静脉分流和几种药物可降低 HVPG)，而不是生存。尚需要研究来证明，肝纤维化重构减少导致的 HVPG 改善可转化为临床结局的改善。因此，FDA 最近关于 NASH 肝硬化临床试验的指导原则没有讨论在这种情况下将 HVPG 作为终点^[38]。需要开发和验证可用于评估 NASH 相关肝硬化患者向死亡及肝脏相关结局进展的非侵入性工具，并确定这些标志物的改善可以预测疾病稳定或改善。这些检测可以包括疾病活动、纤维化、纤维形成、或定量肝功能的标志物。

对于儿童的考虑

儿童疾病的研究对于了解 NAFLD 的自然历史是至关重要的。由于全球肥胖流行，儿童 NAFLD 的患病率很高，估计有 10% 的儿童受到影响^[45]。通常，主要针对年龄大约在 12~14 岁的儿童考虑采取医疗干预。患有 NAFLD 的儿童很大比例在成年后仍有该病。

与儿童 NAFLD 相关的危险因素包括肥胖、男性、青春期、家族一致性、某些代谢综合征^[45,46]。约三分之一的儿童有组织学特征儿童亚型：通常表现为肝脂肪变性更明显，汇管区炎症和纤维化较明显，小叶炎症很少或没有，很少或没有细胞气球样变，没有窦周纤维化^[45,46]。儿童 NAFLD/NASH 领域存在着多重知识空白 (例如自然史、疾病表型、最佳治疗方法)。此外，儿童期组织学亚型 (脂肪变性和纤维化的分区) 可能会影响以后生活中脂肪性肝炎的发展或更快的纤维化进展^[45,46]。

对 NAFLD 儿童患者进行临床研究存在特殊的挑战，包括实际问题考虑（如不能/不愿意服药、恐惧、剂量不确定）、安全问题（例如与研究相关的活检程序、药物毒性），以及关于临床终点获益的不确定性（例如死亡率）（图 2）。多方面知识之间缺乏联系也给研究设计带来了挑战。以什么样的速度进展为可能导致有临床意义结局的肝脏疾病尚不确定。这是目前为这些儿童，尤其是有发展为进展期肝病风险的儿童，进行药物开发面临的最主要障碍。

儿童 NASH 的临床试验障碍

治疗的潜在获益必须与不存在疾病进展风险而使受试者暴露于药物的风险进行权衡。因此，应分析儿童研究中获取的数据，以寻找有肝脏相关结局风险的亚群，从而优先考虑高危患者，并避免低风险患者不必要地暴露于具有潜在毒性的药物。在这个问题达成广泛共识之前，儿童药物研究发展计划应重点关注有最高肝脏相关结局（桥接纤维化或肝硬化）风险的儿童。对于安全性和耐受性已确定的药物，也可以同时考虑 NASH 合并纤维化程度较低的儿童患者，因为要了解一种治疗药物能否减缓向肝硬化的进展（reduced progression to cirrhosis）需要多年的时间。而减缓向肝硬化的进展是目前药物获得完全批准普遍接受的替代终点。

在长期儿童研究中，进入青春期，会涉及到大量的生物学、激素和行为变化，这些变化会影响代谢状态。此外，依从性、对酒精或毒品的尝试会影响试验结果，研究设计和实施应考虑到这些问题。

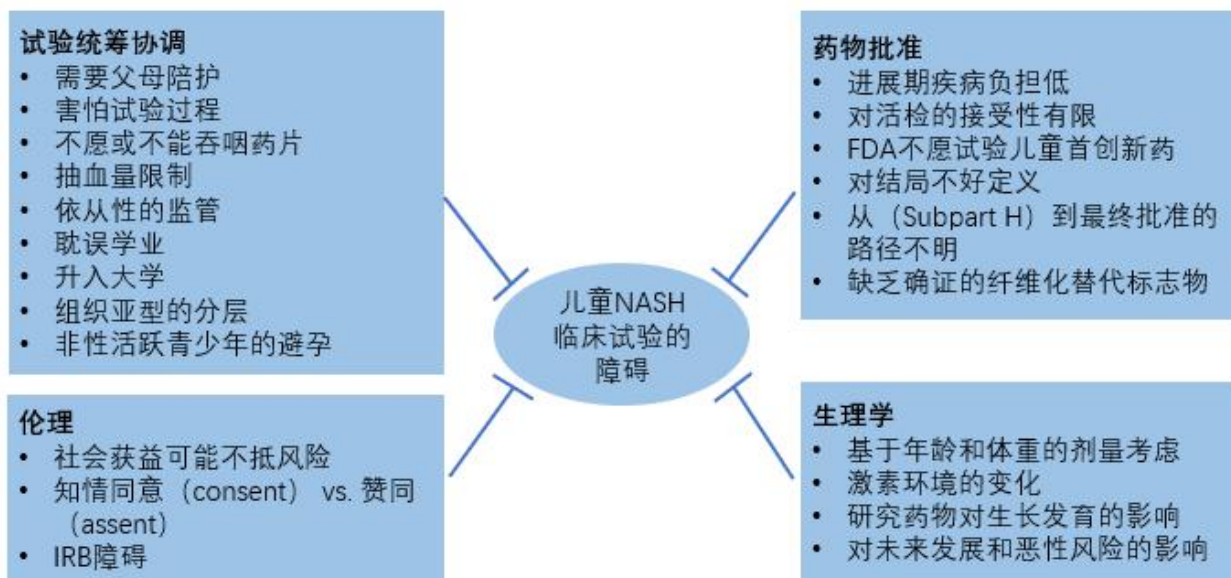


图 2: NASH 患儿药物开发中的潜在障碍。这些障碍包括统筹协调障碍如对程序的恐惧、伦理障碍、监管障碍和生理障碍等。FDA: 美国食品和药物管理局; IRB: internal review board, 内部审查委员会。

译者注: Consent 用于达到给出同意法定年龄的个体(在美国, 通常是 18 岁)。Assent 用于无法给出法律效力 Consent 的个体。对于不能给予 Consent 的儿童或成人, 需要得到他的 Assent, 同时需要父母或合法监护人的 Consent。

尽管两项重要的由美国糖尿病、消化和肾脏疾病国家研究所 NASH 临床研究网络完成的多中心随机对照试验使用预先指定的组织学终点作为主要或次要终点, 但坚持将组织学作为主要终点的较少。鉴于对安全性的关注, 包括对肝活检风险的关注, 与成人患者相比, 使用组织学的替代终点作为试验终点更容易被接受。虽然这些试验主要证明了肝脏组织学替代终点作为儿童试验的可行性, 但确证儿童 NAFLD 进展的替代标记物将进一步提高学术机构和家庭对临床研究的接受程度, 并将进一步降低研究相关风险。虽然可以接受未充分验证的替代指标 (如 ALT、MRI-PDF) 作为试验终点以期将儿童/青少年的操作相关风险降到最低, 但是基于组织学的试验显示, ALT 或其他 NAFLD 严重程度替代指标的变

化与组织学特征之间存在显著的不一致性。鉴于此，在儿童患者中主张进行以组织学为基础的研究证明非劣效性和有效性。

总结

综上所述，我们对NAFLD的自然病程和临床试验对象的选择有了进一步的了解，尽管根据潜在的疾病生物学和疾病进展速度对病人进行精细分组仍然存在很大的知识空白。对并存病和疾病影响因素认知的提高，以及技术的快速进步（例如，成像技术、微生物群、基因分析）支持在可预见的将来实现对NAFLD患者的更细分层。同样，尽管人们已经获得了大量关于疾病发生发展分子机制的数据，但需要识别出可以让大多数病人获益的关键性治疗靶点。关于肝病进展的临床试验，其研究终点需反映干预所针对的发病机制，同时需监测合并的心血管和代谢性疾病以及确保安全性。试验结果透明地公布是至关重要的。终点的选择可能会进一步完善，以反映疾病过程中新出现的数据，以及那些将特异替代指标与有临床意义的结局联系起来的数据。