·指南与规范·

DOI：10.12449/JCH260811

《2026年欧洲肝病学会和美国肝病学会德尔菲共识声明：原发性胆汁性胆管炎替代终点和真实世界证据》摘译

葛启超，蔡晓波，陆伦根

上海交通大学医学院附属第一人民医院消化科，上海 200080

通信作者：陆伦根，lungenlu1965@163.com（ORCID：0000-0002-1533-4068）

摘要：2026年，欧洲肝病学会和美国肝病学会联合发布《原发性胆汁性胆管炎替代终点和真实世界证据德尔菲共识声明》。该共识围绕原发性胆汁性胆管炎新药研发中最关键的3类问题形成16条声明和42条推荐意见：一是肝脏生化指标及非侵入性肝纤维化评估能否作为临床结局的替代终点；二是真实世界数据和真实世界证据如何以符合监管要求的方式补充或支持确证性研究；三是患者报告结局如何标准化用于生活质量、瘙痒和疲乏评价。本文对该共识的背景、方法学框架、核心声明、推荐意见、关键图表进行摘译与专家解读，旨在为国内原发性胆汁性胆管炎诊疗、药物评价、真实世界研究设计和以患者为中心的临床终点建设提供参考。

关键词：原发性胆汁性胆管炎；替代终点；真实世界证据；欧洲；美国；共识

An excerpt of EASL-AASLD Delphi consensus statement on surrogate endpoints and real-world evidence in primary biliary cholangitis (2026)

Ge Qichao, *Cai Xiaobo*, *Lu Lungen*

Department of Gastroenterology, *Shanghai General Hospital*, *Shanghai Jiao Tong University School of Medicine*, *Shanghai 200080*, *China*

Corresponding author: *Lu Lungen*, *lungenlu1965@163.com* (*ORCID*: *0000-0002-1533-4068*)

Abstract: In 2026, the European Association for the Study of the Liver and the American Association for the Study of Liver Diseases jointly published a Delphi consensus statement on surrogate endpoints and real-world evidence in primary biliary cholangitis. This consensus proposes 16 statements and 42 recommendations centering on the following three most critical issues in the development of new drugs for primary biliary cholangitis: whether liver biochemical parameters and noninvasive assessments of liver fibrosis can be used as surrogate endpoints for clinical outcomes; how real-world data and real-world evidence can complement or support confirmatory studies in a manner compliant with regulatory requirements; how patient-reported outcomes can be standardized for the assessment of quality of life, pruritus, and fatigue. This article provides an excerpt and expert interpretation of the background, methodological framework, core statements, recommendations, and key figures and tables of this consensus, in order to provide a reference for the diagnosis and treatment of primary biliary cholangitis, drug evaluation, real-world study design, and the development of patient-centered clinical endpoints in China.

Key words: Primary Biliary Cholangitis; Surrogate Endpoint; Real-World Evidence; Europe; United States; Consensus

原发性胆汁性胆管炎（primary biliary cholangitis，PBC）是一类以慢性胆汁淤积、小胆管免疫介导性损伤、进行性纤维化及生活质量受损为特征的自身免疫性肝病。熊去氧胆酸（ursodeoxycholic acid，UDCA）改善了PBC自然史，但仍有相当比例患者不能获得充分生化应答，或因持续胆汁淤积、纤维化进展、门静脉高压并发症及肝移植/死亡风险而需要二线治疗。近年来奥贝胆酸、seladelpar和elafibranor等药物陆续基于碱性磷酸酶（alkaline

phosphatase，ALP）与总胆红素改善获得条件性或加速批准，苯扎贝特和非诺贝特也在部分国家和地区被广泛超说明书应用。上述进展推动了PBC治疗格局从“单一UDCA时代”进入“风险分层、多靶点和症状管理并重”的阶段。

然而，PBC新药研发的核心瓶颈并未消失。PBC进展缓慢，肝功能失代偿、肝移植和死亡等硬终点发生率低，若以这些结局作为Ⅲ期随机对照试验（randomized controlled trial，RCT）主要终点，研究通常需要多年随访和大量样本，现实可行性差。药物一旦获得条件性批准并进入临床，安慰剂随机、长期维持盲法、患者留组和交叉用药控制也会显著困难。COBALT研究的受挫提示，传统“以硬终点确证”的监管路径，在PBC这类患病率较低、终点事件发生率较低的慢性肝病研究中，同时面临伦理层面与操作层面的双重挑战。

本部共识[1]的重要性在于，其并非单纯复述PBC治疗建议，而是直面“如何证明药物真正改变疾病结局、如何用真实世界证据（real world evidence，RWE）弥补RCT不足、如何将患者感受纳入监管可接受终点”3个基础性问题。其价值不仅限于PBC，也可为原发性硬化性胆管炎、自身免疫性肝炎及其他罕见胆汁淤积性肝病的终点选择和研究设计提供方法学模板。本文在摘译原文要点的基础上，结合国内临床研究实际，对关键内容进行结构化整理。

1 共识形成方法与总体框架

该共识由欧洲肝病学会（European Association for the Study of the Liver，EASL）和美国肝病学会（American Association for the Study of the Liver Diseases，AASLD）发起，并与拉丁美洲肝脏研究协会及非洲肝病学会协作。专家组最初包含62名成员，涵盖临床专家、方法学专家、监管机构代表、工业界代表和患者代表；最终德尔菲投票专家来自18个国家，详细特征见表1。工作组围绕3个主题开展：PBC临床试验替代终点、PBC真实世界数据（real world data，RWD）/RWE、PBC患者报告结局。文献检索遵循PRISMA原则，并采用牛津循证医学中心标准评估证据等级。

德尔菲过程设置两轮投票。第一轮针对16条声明和42条推荐意见投票，回应率为88%；第二轮对需要修改的4条推荐意见和1条声明再次投票，回应率为60%。监管机构代表参与讨论但不参与投票，工业界每个组织仅允许1票，以降低利益不平衡。该设计体现了本共识的一个鲜明特点：从一开始即将临床、患者、监管和研发各方置于同一讨论框架内，目标不是形成传统诊疗流程，而是形成可被临床试验和监管评价采用的方法学路线图。

表1 德尔菲专家组特征

Table 1 Delphi panel characteristics

项目	数值
专家组总人数	53
临床专家	34（64%）
工业界代表	8（15%）
患者代表	8（15%）
监管代表[1)]	3（6%）
男性/女性	32（60%）/21（40%）
欧洲/北美/亚太/南美/非洲	29（55%）/18（33%）/3（6%）/2（4%）/1（2%）
覆盖国家	18
第1轮德尔菲投票回应者	44（88%）[2)]
第2轮德尔菲投票回应者	30（60%）[2)]

注：1）监管代表参与讨论但不投票；2）分母不包括监管代表，且每个工业组织仅计1票。

2 PBC临床试验中的替代终点

共识强调，Ⅲ期PBC RCT的主要理想终点应当是患者相关且具有临床意义的结局，即能够反映患者“感觉、功能或生存”的改变。但由于PBC硬终点发生缓慢，替代终点在药物研发中不可或缺。当前任何单一PBC生物标志物尚未完全达到监管机构意义上的“已验证替代终点”，但ALP、总胆红素、白蛋白、血小板计数及振动控制瞬时弹性成像（vibration controlled transient elastography，VCTE）测量的肝硬度值（liver stiffness measurement，LSM）已经具备较一致的预后关联证据。

2.1 肝脏生化指标：ALP和总胆红素仍是核心 ALP与总胆红素是当前PBC研究最为充分的生化指标。一项纳入1995—2025年共计69项研究的系统综述显示，样本量中位数为540例，随访时长为2.5～10年。无论在未经治疗的自然史队列、UDCA治疗队列，抑或是奥贝胆酸、苯扎贝特等二线治疗队列中，患者治疗12个月时的ALP与总胆红素水平，均与肝功能失代偿、肝移植及死亡风险显著相关。现有最大规模国际队列研究结果提示，当ALP>2倍正常值上限（upper limit of normal，ULN）时，患者发生不良临床结局的风险会出现升高；若总胆红素>1×ULN，上述风险升高则更为显著。

共识认为，ALP和总胆红素可以继续作为PBC条件性/加速批准试验的关键终点，但未来应逐步提高终点严谨性。完全生化应答（ALP和总胆红素均正常）具有更强预后价值；完全生化应答定义为ALP<1.5～1.67×ULN且总胆红素正常。需要注意的是，天冬氨酸氨基转移酶、丙氨酸氨基转移酶和γ-谷氨酰转移酶虽然常规检测，但目前证据不足以单独判定其对长期结局的独立影

响。白蛋白和血小板计数反映肝合成功能及门静脉高压/进展性肝病风险，具有预后价值，可作为风险分层和复合终点构建的组成部分。

推荐意见1：在基于生化入组标准、持续时间至少12个月的Ⅲ期PBC临床试验中，完全生化应答（ALP和总胆红素均正常）应被视为理想终点；完全生化应答（ALP<1.5～1.67×ULN且总胆红素正常）可作为可接受终点（证据等级2，强推荐，一致率90%）。

共识声明要点：UDCA治疗12个月后ALP不同阈值（正常、1.5×ULN、1.67×ULN、2×ULN）和总胆红素不同阈值（0.6×ULN、1×ULN、2×ULN）均与临床结局相关；二线治疗12个月后ALP和总胆红素也与临床结局相关；UDCA治疗6～12个月后白蛋白和血小板计数具有预后意义；UDCA治疗6～12个月或之后任意时间未获得充分生化应答提示不良结局风险增加。

2.2 非侵入性肝纤维化评估：LSM是最具前景的疾病修饰终点 组织学进展理论上可作为PBC疾病修饰的替代终点，但重复肝脏活组织检查有创、患者接受度低，且受取样误差和观察者差异影响。VCTE测量的LSM是PBC中验证最充分的非侵入性纤维化指标。现有研究提示，基线或治疗中LSM>10 kPa与不良临床结局相关，且这种关联独立于生化指标；1年内LSM升高≥20%提示结局风险增加，而下降≥20%提示风险改善。

该共识同时指出，LSM并非理想的替代终点。胆汁淤积、炎症反应、操作者技术水平、患者肥胖因素，以及不同中心间的测量一致性均可能对检测结果产生影响；与此同时，前瞻性验证LSM动态变化与硬终点之间因果关联的证据仍较为有限。增强肝纤维化评分在基线风险预测方面已积累一定研究证据，但治疗过程中该指标的纵向变化规律及其与生化应答的关联尚不明确。基于上述分析，该共识建议将LSM与生化应答组合，作为随访周期足够长的临床试验的复合终点。针对早期疾病的相关研究，若以纤维化稳定或下降作为研究终点，随访周期至少应设置为24个月（图1）。

推荐意见2：在试验周期适当的PBC Ⅲ期临床试验中，可将VCTE测量的LSM作为临床结局的潜在替代指标（证据等级3，弱推荐，一致率92%）。

推荐意见3：在试验周期适当（早期疾病至少24个月）的PBC Ⅲ期临床试验中，应考虑将生化应答联合LSM下降或稳定作为复合终点（证据等级2，强推荐，一致率86%）。

推荐意见4：可将生化应答不足、LSM进展和/或肝功能失代偿组成联合终点，用于判定治疗失败（证据等级3，弱推荐，一致率98%）。

3 RWD和RWE

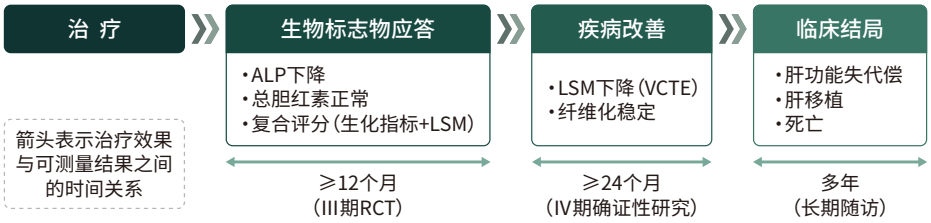
本共识对RWD/RWE的定位作出明确界定：RWD/RWE除可用于描述疾病自然史、识别风险因素、优化样本量估计与开展安全性监测外，更可在RCT无法开展的场景下，为药物有效性与安全性的因果推断提供支撑。用于监管决策的RWE与常规观察性研究存在差异，其开展需满足以下前提：选用具备“适用性”的数据源、预先制定研究方案、数据来源公开透明、分析流程具备可重复性，且对研究中存在的偏倚与混杂因素进行充分识别与控制。

3.1 研究设计：早期监管沟通与外部对照

推荐意见5：拟用于支持药物批准或适应证扩展的研究，应尽早并持续与监管机构沟通（证据等级5，强推荐，一致率100%）。

推荐意见6：当RCT无法开展时，采用单臂试验联合经由严格定义的RWD构建的外部对照，可作为支持药物安全性与有效性评价、监管级别证据的可接受研究设计（证据等级3，强推荐，一致率100%）。

专家解读：上述两项推荐意见对于中国PBC研究具有直接启示。国内PBC患者总体数量有限，多中心长期安慰剂对照研究面临招募和伦理压力。若要使本土真实世界研究进入药物评价链条，不能停留在“回顾性病例汇总”层面，而应在研究启动前即明确目标试验模拟、外部



注：PBC，原发性胆汁性胆管炎；ALP，碱性磷酸酶；LSM，肝硬度值；VCTE，振动控制瞬时弹性成像。

图1 PBC临床试验替代终点框架

Figure 1 Surrogate endpoint framework for PBC clinical trials

对照构建、入排标准、暴露定义、终点定义、缺失数据处理、敏感性分析和监管沟通路径。

3.2 数据源：注册队列、电子健康记录、医保/理赔数据库和分布式网络互补

推荐意见7：PBC的RWD研究应根据研究目的和数据适用性选择合适数据源，可包括疾病注册系统、电子健康记录、理赔/医保数据库和分布式数据网络（证据等级5，强推荐，一致率98%）。

推荐意见8：不同数据源的终点定义应该统一，并与监管标准保持一致（证据等级5，强推荐，一致率100%）。

推荐意见9：拟用于生成监管级证据的新兴或既有数据源，应通过结构化方法解决关键知识缺口，以提高数据质量和使用价值（证据等级5，强推荐，一致率100%）。

疾病特异性注册队列可提供粒度最精细的治疗暴露、实验室动态变化、LSM、患者报告结局以及长期结局数据，均为验证风险模型与替代终点的核心基础；电子健康记录更贴合常规临床诊疗场景，可记录合并症、处方路径与安全性事件；医保或理赔数据库覆盖范围广，便于识别住院、肝移植、医疗费用等硬终点，但通常缺少实验室检查结果与患者报告结局相关细节；分布式网络有助于开展跨区域数据联邦分析，同时可实现隐私保护。国内应当推动PBC专病注册系统与医院信息系统、医保数据及患者随访平台建立标准化互联互通。

3.3 偏倚控制：预先指定、内部验证和透明报告

推荐意见10：所有PBC真实世界研究均应预先指定预期偏倚（表2）（证据等级5，强推荐，一致率100%）。

推荐意见11：使用RWD的PBC研究应预先规定并实施偏倚控制策略（证据等级5，强推荐，一致率100%）。

推荐意见12：电子健康记录和理赔数据库应通过与临床参考标准对照或人工病历核查进行内部验证，以确保研究人群、暴露、结局和协变量识别准确（证据等级2，强推荐，一致率100%）。

推荐意见13：若采用代币化技术链接去标识化患者数据，应说明代币选择理由，并量化链接错误风险，如错误匹配和漏匹配（证据等级5，强推荐，一致率100%）。

推荐意见14：当RWD分析出现非预期或不确定结果时，应当鼓励开展事后偏倚来源调查，并对调查过程与结果进行透明报告，以为后续研究及监管决策提供支撑（证据等级5，强推荐，一致率100%）。

4 患者报告结局：从“可选指标”转向“核心终点”

PBC患者常长期承受瘙痒、疲乏、睡眠障碍、认知功能受损、情绪/社会功能受限等负担。UDCA和部分二线药物可以改善生化指标，但对症状影响有限。共识明确提出，患者报告结局不是附属指标，而是PBC研究的基础性终点；只有系统纳入患者报告结局，才能真正反映患者“感觉”和“功能”的改善，也才能为药物标签、卫生技术评估和临床决策提供患者中心证据（表3）。

患者报告结局使用的关键在于标准化。研究应在多个预先指定时间点收集患者报告结局，并与治疗反应

表2 PBC真实世界研究常见偏倚及缓解策略

Table 2 Biases relevant to real-world data studies in PBC and recommended mitigation strategies

偏倚类型	简要解释	统计/设计缓解策略	PBC研究示例
选择偏倚	研究人群可能无法代表普通PBC人群	人群加权、敏感性分析、扩大数据源	登记的数据通常代表依从性较高的患者；国际疾病分类编码敏感性/特异性有限
适应证混杂/渠道偏倚	病情重者更可能接受特定治疗	倾向评分匹配或加权、多变量校正	进展性疾病患者更可能启动二线治疗；医保可及性影响用药
永恒时间偏倚	入组至治疗开始之间未正确计入的时间造成生存优势假象	时间依赖Cox模型、统一索引日期、严格时间对齐	若以治疗启动日作为零点，可能人为产生生存偏倚
信息偏倚	数据采集不准确或不一致导致暴露/结局判定错误	验证研究、标准化采集流程、算法与病历核查比对	肝硬化诊断算法需验证；移植等待名单可能比实际移植更接近真实疾病进展
误分类偏倚	暴露或结局分类错误扭曲关联	验证算法、敏感性分析、独立终点判定委员会	需比较意向性分析与实际治疗分析；不查阅病历时不良事件捕捉困难
随访时间不一致	测量频率差异影响结局发现	时间变化协变量建模、按随访强度校正、主动对照设计	症状重者就诊更频繁；用药者因续方产生规律随访
症状/患者报告结局记录偏倚	主观症状可能漏报或记录不一致	结构化症状工具、标准化患者报告结局量表	患者报告结局可影响治疗启动和停药；顽固瘙痒可独立驱动移植登记
缺失数据偏倚	信息缺失若非随机可导致估计偏倚	多重插补、针对非随机缺失的敏感性分析	病情过轻或过重均可能增加缺失；无症状稳定患者随访间隔长

注：PBC，原发性胆汁性胆管炎。

表3 PBC研究中理想的患者报告结局工具特征

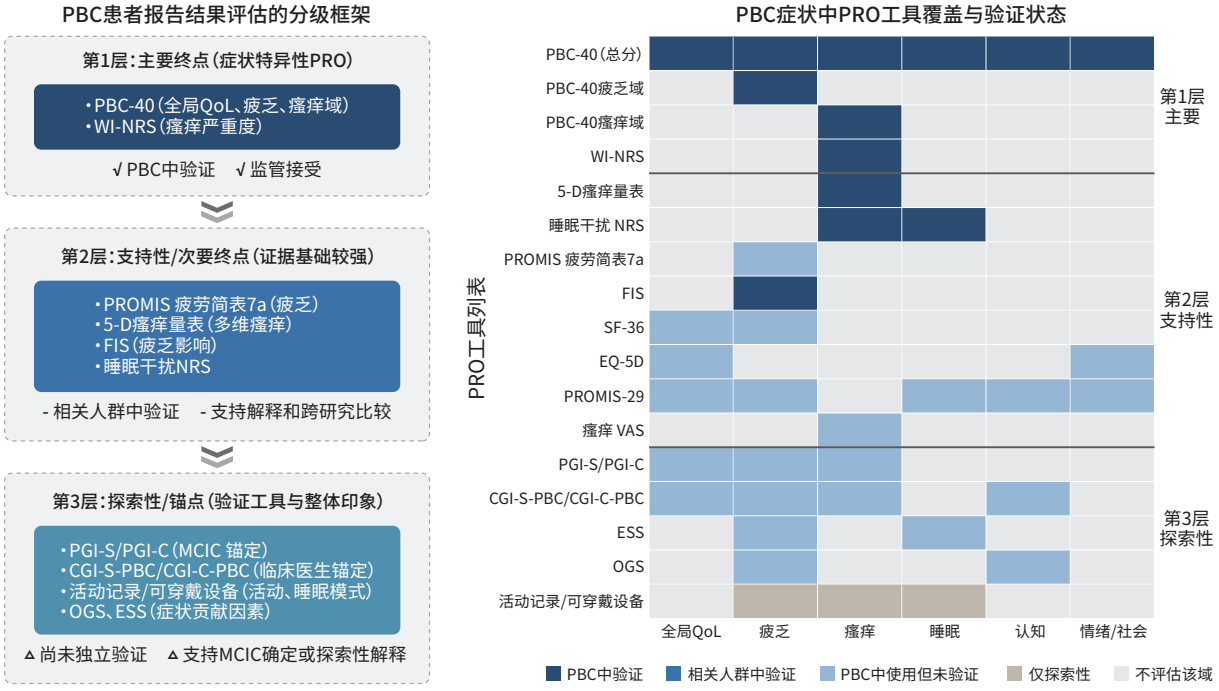
Table 3 Ideal characteristics of patient- reported outcomes instruments for use in PBC

特征	说明	理由
经PBC人群验证	可信性、有效性和反应性已在PBC患者中验证	确保在特定疾病背景下能够准确测量预期的指标
患者来源内容	条目内容基于PBC患者资料而拟定	捕捉患者真实经历和语言
覆盖PBC相关症状域	涵盖主要症状领域：疲劳、瘙痒、认知功能、情绪状态以及社会参与	反映生化指标以外的多维疾病负担
建立临床重要变化最小阈值	使用锚定方法定义最小临床重要变化	帮助判断变化对患者是否有意义
对变化敏感	能灵敏检测治疗前后的变化	用于临床试验疗效评价
监管熟悉度	已被FDA/EMA接受或审阅用于临床试验	有利于监管申报和标签声明
合适的回忆期	回忆期应与症状波动和研究设计匹配	减少回忆偏倚，准确捕捉如瘙痒等波动症状
适合重复使用	受试者负担低，可多次施测	支持长期纵向评价，且不会造成患者过度疲劳
适合多语种应用	具备多语言的翻译版本	保障国际/多地区研究的跨文化可对比性
兼容电子采集	经患者报告结局的电子平台验证	便于实时采集数据，并与电子健康记录/注册系统兼容

注：PBC，原发性胆汁性胆管炎；FDA，美国食品药品监督管理局；EMA，欧洲药品管理局。

预期轨迹一致；每个量表的回忆期和评分方法应在方案或统计分析计划中预先定义；最小临床重要变化阈值应预先建立；患者总体印象严重程度量表（patient global impression of severity，PGI-S）/患者总体印象变化量表（patient global impression of change，PGI-C）和PBC临床总体印象严重程度量表（clinical global impression of severity for PBC，CGI-S-PBC）/PBC临床总体印象变化量表（clinical global impression of change for PBC，CGI-C-PBC）等整体印象工具可作为锚点帮助解释患者报告结局变化；同一研究中应尽量使用量表验证时采用的电子或纸质形式，并保持所有受试者和随访时间点一致（图2）。

推荐意见15：患者报告结局应在多个预先指定时间点采



注：PBC，原发性胆汁性胆管炎；PRO，患者报告结局；PBC-40，原发性胆汁性胆管炎40项问卷；QoL，生活质量；WI-NRS，最严重瘙痒数字评定量表；PROMIS，患者报告结局测量信息系统；FIS，疲劳影响量表；NRS，数字评定量表；PGI-S，患者总体印象严重程度量表；PGI-C，患者总体印象变化量表；CGI-S-PBC，PBC临床总体印象严重程度量表；CGI-C-PBC，PBC临床总体印象变化量表；OGS，直立性分级量表；ESS，Epworth嗜睡量表；MCIC，最小临床重要变化；SF-36，36项简明健康调查量表；EQ-5D，欧洲五维健康量表；VAS，视觉模拟量表。

图2 PBC中患者报告结局分层评估框架

Figure 2 Tiered framework for patient-reported outcome assessment in PBC

集，并与试验设计和预期治疗反应轨迹相匹配（证据等级2，强推荐，一致率98%）。

推荐意见16：研究应清楚界定每个工具的回忆期和评分方法；在临床试验中，应在研究方案或统计分析计划中预先说明（证据等级2，强推荐，一致率100%）。

推荐意见17：应预先建立最小临床重要变化阈值，为患者个体水平及群体水平的治疗效应解释提供支撑（证据等级2，强推荐，一致率100%）。

推荐意见18：PGI-S/PGI-C，以及适用场景下的CGI-S-PBC/CGI-C-PBC，可用于辅助阐释患者报告结局的变化，尤其适合作为构建最小临床重要变化阈值的锚定工具（证据等级5，弱推荐，一致率98%）。

推荐意见19：患者报告结局理想情况下应采用其验证时使用的给药形式（电子或纸质）；无论选择何种形式，研究内应保持一致（证据等级5，强推荐，一致率100%）。

推荐意见20：建议对患者报告结局进行及时且相对频繁的采集，以降低回忆偏倚，并准确捕捉症状波动情况（证据等级5，强推荐，一致率98%）。

推荐意见21：活动记录、体位/自主神经测试等客观工具可作为探索性或支持性指标，但在进一步验证前不能替代患者报告结局作为监管终点（证据等级5，弱推荐，一致率100%）。

推荐意见22：临床研究中应对可能影响症状解读的合并症与症状调节机制，如甲状腺疾病、抑郁、睡眠障碍、自主神经功能异常等，进行测量及预设分析（证据等级5，强推荐，一致率100%）。

4.1 总体生活质量：原发性胆汁性胆管炎40项问卷（primary biliary cholangitis-40，PBC-40）应作为核心疾病特异性工具（表4）

推荐意见23：PBC-40应被视为评估PBC总体生活质量的主要疾病特异性工具（证据等级2，强推荐，一致率96%）。

推荐意见24：欧洲五维健康量表（EuroQol 5-dimensions questionnaire，EQ-5D）、36项简明健康调查量表（short form-36 health survey，SF-36）和患者报告结局测量信息系统29项问卷（patient-reported outcomes measurement information system-29，PROMIS-29）不应单独用于评估PBC生活质量（证据等级5，强推荐，一致率100%）。

推荐意见25：PGI-S/PGI-C等患者整体印象工具可用于建立有意义的患者内变化阈值（证据等级5，弱推荐，一致率100%）。

PBC-40是疾病特异性量表，覆盖总体生活质量、疲乏、瘙痒、认知、情绪和社会功能等维度，最能反映PBC患者综合负担。EQ-5D、SF-36和PROMIS-29具有跨疾病可比性和卫生经济学价值，但对PBC症状变化的敏感度不足，适合作为支持性工具，而非单独主终点。

4.2 瘙痒：最严重瘙痒数字评定量表（worst itch numerical rating scale，WI-NRS）是主要量表，需兼顾强度与影响（表4）

推荐意见26：WI-NRS推荐作为PBC临床试验中评估瘙痒严重度的主要工具（证据等级2，强推荐，一致率98%）。

推荐意见27：瘙痒试验中，目前推荐以研究期间治疗组与安慰剂组WI-NRS较基线平均变化差异作为主要终点（证据等级2，强推荐，一致率94%）。

推荐意见28：推荐将WI-NRS（0~10分量表）较基线的下降幅度≥3分，作为PBC临床试验中评估患者个体临床获益的主要阈值（证据等级2，强推荐，一致率94%）。

推荐意见29：PGI-S和PGI-C等整体印象工具可辅助解读患者个体内具有临床意义的变化（证据等级5，弱推荐，一致率96%）。

推荐意见30：可穿戴技术（如活动记录）可作为补充患者报告结局的探索性数据，但尚未验证为监管试验终点（证据等级5，弱推荐，一致率100%）。

推荐意见31：视觉模拟量表可用于支持性或真实世界研究场景（证据等级5，强推荐，一致率98%）。

推荐意见32：能够全面反映瘙痒更广泛影响的多维评估工具，如5-D瘙痒量表和PBC-40瘙痒域，应被考虑作为次要终点或复合终点组成部分（证据等级5，强推荐，一致率96%）。

推荐意见33：用于评估瘙痒相关后果的工具，例如睡眠干扰数字评定量表或PROMIS睡眠障碍量表，可作为中重度瘙痒患者临床试验的次要终点（证据等级5，弱推荐，一致率96%）。

推荐意见34：EQ-5D、SF-36等生活质量量表可用于阐明治疗效应的背景信息，但此类量表针对靶向症状干预的变化敏感性较低，因此不宜作为瘙痒特异性临床试验的主要终点或次要终点（证据等级5，弱推荐，一致率100%）。

专家解读：PBC相关瘙痒存在显著的昼夜波动特征，可对患者睡眠产生影响，并引发相应心理社会不良后果，若仅对瘙痒强度进行评估，可能低估干预方案的实际治疗获益。在国内以瘙痒为核心适应证开展相关临床试验时，除WI-NRS外，还应纳入PBC-40瘙痒域、5-D瘙痒量表与睡眠干扰量表，同时预先设置瘙痒评分下降2分、3分、4分等多阈值的敏感性分析，以此兼顾监管要求的可接受性与患者的个体异质性。

表4 PBC患者报告预后指标的验证状态总结

Table 4 Summary of validation status for patient-reported outcome instruments in PBC

工具	评估域	PBC中验证	PBC中提出最小临床重要变化	临床研究推荐用途	备注
PBC-40	总体生活质量、疲乏、瘙痒、认知、情绪、社会功能	是	否	试验主要终点	疾病特异性，内容来自患者
PBC-40疲乏域	疲乏严重度和影响	是	否	次要/共同主要终点	核心疾病特异性疲乏工具
PBC-40瘙痒域	瘙痒影响	是	否	次要终点	对瘙痒强度反映不足
WI-NRS	瘙痒严重度	是	是（下降≥3分）	瘙痒试验主要终点	FDA/EMA接受
5-D瘙痒量表	持续时间、程度、方向、伤残、分布等多维瘙痒	否[1)]	否	次要终点	慢性瘙痒中验证，PBC队列使用过
瘙痒视觉模拟量表	瘙痒严重度	否	否	支持性/RWE	简单但标准化不足
睡眠干扰NRS	瘙痒导致睡眠障碍	否[2)]	否	次要终点	在瘙痒人群中验证
PROMIS疲劳简表7a	一般疲乏	否[3)]	否	互补工具	心理测量学强，便于跨疾病比较
疲乏影响量表	认知、躯体和社会维度疲乏	否[1)]	否	支持性工具	在肝病中验证
EQ-5D	健康效用	否[2)]	不适用	经济学建模	非疾病特异，HTA认可
SF-36	一般健康相关生活质量	是	否	次要终点	在肝病中验证，缺乏PBC症状特异性
PROMIS-29	功能、情绪、睡眠、疼痛、疲乏等	否[2)]	否	互补工具	覆盖广，可推导效用
PGI-S/PGI-C	患者整体严重度/变化印象	否[2)]	锚定功能	最小临床重要变化锚点	支持解释，不宜单独作终点
CGI-S-PBC/CGI-C-PBC	临床医生评价严重度/变化	否[3)]	锚定功能	最小临床重要变化锚点	疾病特异性，正在验证
Epworth嗜睡量表	日间嗜睡	否	否	支持性	反映疲乏相关因素
体位分级量表	自主神经症状	否	否	支持性	有助于疲乏表型刻画
活动记录/可穿戴设备	客观活动/睡眠	否	否	仅探索性	尚未验证为监管终点

注：1)在其他肝病或慢性瘙痒人群中验证；2)PBC试验中使用但未在PBC中特异验证；3)正在PBC中验证。PBC，原发性胆汁性胆管炎；PBC-40，原发性胆汁性胆管炎40项问卷；WI-NRS，最严重瘙痒数字评定量表；NRS，数字评定量表；PROMIS，患者报告结局测量信息系统；EQ-5D，欧洲五维健康量表；SF-36，36项简明健康调查量表；PGI-S，患者总体印象严重程度量表；PGI-C，患者总体印象变化量表；CGI-S-PBC，PBC临床总体印象严重程度量表；CGI-C-PBC，PBC临床总体印象变化量表；HTA，卫生技术评估；FDA，美国食品药品监督管理局；EMA，欧洲药品管理局。

4.3 疲乏：应以疾病特异性患者报告结局为核心，客观工具仅可用于探索性研究（表4）

推荐意见35：疲乏评估需采用可同时捕捉其中枢性与外周性构成特征的患者报告结局工具，并充分认可PBC相关疲乏存在多因素致病与个体化表现的特征（证据等级2，强推荐，一致率100%）。

推荐意见36：PBC-40疲乏域应作为PBC疲乏的主要疾病特异性评估工具（证据等级2，强推荐，一致率96%）。

推荐意见37：PROMIS疲劳简表7a应作为互补疲乏量表纳入（证据等级5，强推荐，一致率98%）。

推荐意见38：疲乏影响量表及其修订版可作为支持性工具（证据等级5，弱推荐，一致率100%）。

推荐意见39：患者和临床医生整体印象工具（PGI-S、PGI-C、CGI-S/C-PBC等）可用于辅助阐释患者个体内的有意义变化，但不得替代患者报告结局作为主要结局指标（证据等级5，强推荐，一致率100%）。

推荐意见40：其他非PBC特异性工具，如医院焦虑抑郁量表、Epworth嗜睡量表和体位分级量表，可补充疲乏评估（证据等级5，弱推荐，一致率98%）。

推荐意见41：活动记录或可穿戴传感器等客观工具可作为PBC探索性或次要终点（证据等级5，弱推荐，一致率98%）。

疲乏是最常见且最困扰PBC患者的症状之一，40%～80%的患者受其影响。疲乏程度与疾病分期并不稳定相关，也可在生化应答后持续存在。其机制可能涉及中枢神经调节、外周肌肉功能、睡眠障碍、情绪障碍和自主神经功能异常。因此，疲乏既是症状终点，也是潜在治疗靶点。国内研究需要避免将疲乏简单归因于“肝功能异常”，应系统排查甲状腺疾病、贫血、抑郁、睡眠障碍和药物影响，并在统计分析中预设校正。

5 对中国临床工作者和研究者的启示

第一，应建立“风险分层-终点分层-治疗分层”一体化思路。国内PBC临床实践中，ALP与总胆红素仍应作为疗效监测和二线治疗决策的核心指标，但不宜仅以“是否下降”笼统判断疗效。对于临床研究，建议预先区分完全生化应答和充分生化应答，并同步采集白蛋白、血小板计数、LSM和症状评分，以便构建生化-纤维化-患者体验三维复合终点。

第二，应将VCTE-LSM纳入PBC长期随访和研究数据库的标准变量。国内不同中心之间的VCTE设备、操作者和报告格式差异较大，若未来用于多中心研究终点，需建立统一质量控制流程，包括空腹状态、探头选择、有效测量次数、统计学描述指标（四分位间距/中位数）、胆汁淤积和炎症干扰因素记录，以及重复测量时间窗。对于早期PBC，LSM作为疾病修饰指标需要足够随访时间，不能用短期变化过度解释抗纤维化疗效。

第三，应推动中国PBC真实世界研究从“病例登记”升级为“监管级证据平台”。这要求数据模型尽量与国际接轨：明确诊断算法、抗线粒体抗体和PBC特异性抗核抗体状态、UDCA剂量与依从性、二线用药、实验室动态、LSM、肝失代偿、移植登记、死亡、妊娠和合并自身免疫病等变量；同时纳入患者报告结局中文验证量表、随访时间窗和数据质量指标。只有这样，未来本土药物或适应证扩展研究才可能利用RWD构建可解释的外部对照。

第四，应当重视患者报告结局的本土化转译与临床落地推广。PBC患者常因“肝功能尚可”而症状被低估，尤其是瘙痒、疲乏和睡眠障碍。建议国内专家组优先推动PBC-40、WI-NRS、PROMIS疲劳简表7a、5-D瘙痒量表和PGI/CGI工具的中文验证、电子化采集和最小临床重要变化阈值研究。门诊实践中可先采用简化组合：WI-NRS评估瘙痒强度，PBC-40或其关键域评估疾病特异性生活质量，焦虑抑郁量表/Epworth嗜睡量表/体位分级量表用于疲乏鉴别和表型刻画。

第五，应在研究设计中预先处理中国情境下的特有偏倚。例如，不同地区二线药物可及性、医保报销政策、三级医院转诊倾向、中医或保健品使用、实验室参考区间差异、随访频率差异和患者对瘙痒/疲乏的表达习惯，均可能影响RWD因果推断。研究方案应预设倾向评分、限制/匹配、中心分层、阴性对照、缺失数据多重插补和敏感性分析，并通过人工病历核查验证关键变量。

第六，应构建适合中国PBC新药研发的复合终点路径。对于以抗胆汁淤积或抗纤维化为主要机制的药物，可采用12个月完全/充分生化应答作为早期主要终点，并在24个月或更长随访中验证LSM稳定/下降和肝事件减少。对于以瘙痒或疲乏为主要适应证的药物，应将WI-NRS或PBC-40疲乏域等患者报告结局作为主要终点，生化指标则作为安全性和疾病背景指标。对于综合疾病修饰药物，推荐探索“生化应答+LSM稳定/下降+患者报告结局改善”的层级式复合终点。

6 结语

2026年EASL-AASLD PBC替代终点和真实世界证据德尔菲共识的核心贡献，是在传统RCT与现实临床之间搭建了一条更可执行、更符合伦理、也更以患者为中心的证据生成路径。共识并未宣称ALP、总胆红素或LSM已经完成全部替代终点验证，而是明确指出其强预后价值、可用于当前临床试验与条件性批准框架，同时要求通过前瞻性研究和监管级RWE继续完成验证闭环。

对于中国PBC领域，真正需要转化的是其方法学精神：以严谨数据源支持真实世界因果推断，以复合终点反映疾病生物学进展，以标准化患者报告结局体现患者真实获益。未来国内多中心协作若能围绕统一数据标准、中文患者报告结局验证、LSM质量控制和长期肝事件随访建立平台，不仅将提升PBC患者管理质量，也可能在罕见胆汁淤积性肝病药物研发中形成具有国际影响力的中国证据。

利益冲突声明：本文不存在任何利益冲突。

作者贡献声明：葛启超负责文献翻译并撰写文章；蔡晓波和陆伦根负责指导文章撰写、审校并最后定稿。

参考文献：

[1] European Association for the Study of the Liver, American Association for the Study of Liver Diseases. EASL-AASLD Delphi consensus statement on surrogate endpoints and real-world evidence in primary biliary cholangitis[J]. J Hepatol, 2026. DOI: 10.1016/j.jhep.2026.03.038. [Epub ahead of print]

收稿日期：2026-06-15；**录用日期：**2026-07-09

本文编辑：王莹

引证本文：葛启超, 蔡晓波, 陆伦根. 《2026年欧洲肝病学会和美国肝病学会德尔菲共识声明: 原发性胆汁性胆管炎替代终点和真实世界证据》摘译[J]. 临床肝胆病杂志, 2026, 42(8): 1807-1814.
Ge Qichao, Cai Xiaobo, Lu Lungen. An excerpt of EASL-AASLD Delphi consensus statement on surrogate endpoints and real-world evidence in primary biliary cholangitis (2026)[J]. J Clin Hepatol, 2026, 42(8): 1807-1814.