

临床指南和专家共识制定中循证医学 证据等级和推荐强度确定方法

王瑞平 李斌

(上海市皮肤病医院临床研究与创新转化中心 上海 200443)

摘要 近年来，随着生物医学技术进步和新药研发的迅猛发展，临床诊疗技术更新和进展十分迅速。为更好地推广新技术、新药物和新方法在临床诊疗过程中的应用，国内外各专业学会（协会）不断制定疾病的临床诊疗指南和专家共识，以指导临床医师更合理地应用最新的诊疗技术和药物。但在制定临床诊疗指南和专家共识时，编写团队经常面临的一个问题是如何确定纳入指南或共识的“药物”“技术”“方法”的循证医学证据等级和专家推荐强度。因此，本文从文献检索入手，基于检索到的不同文献类别详细阐释如何开展文献质量评价，并据此给出循证医学证据等级和专家推荐强度参考标准和确定方法，以供大家开展临床诊疗指南和专家共识制定工作时参考。

关键词 临床诊疗指南 专家共识 证据等级 推荐强度 德尔菲法

中图分类号 : R-3 **文献标志码** : C **文章编号** : 1006-1533(2025)08-0026-06

引用本文 王瑞平, 李斌. 临床指南和专家共识制定中循证医学证据等级和推荐强度确定方法 [J]. 上海医药, 2025, 46(8): 26-31.

Methods for determining the levels of evidence and strength of recommendations in the development of clinical guidelines and expert consensus

WANG Ruiping, LI Bin

(Clinical Research & Innovation Center, Shanghai Skin Disease Hospital, Shanghai 200443, China)

ABSTRACT In recent years, with the rapid advancement of biomedical technologies and the development of new drugs, clinical diagnostic and therapeutic techniques have been continuously updated and improved. To better promote the application of new technologies, drugs and methods in clinical practice, various professional societies and associations worldwide have been formulating clinical practice guidelines and expert consensus statements to guide clinicians in the rational use of the latest diagnostic and therapeutic approaches. However, during the development of clinical practice guidelines and expert consensus, a common challenge faced by writing groups is how to determine the levels of evidence for the included “drugs”, “technologies” and “methods”, as well as the strength of expert recommendations. Therefore, this article starts with literature retrieval, elaborates on how to conduct literature quality appraisal based on different categories of retrieved studies, and subsequently provides reference standards and methods for determining levels of evidence and the strength of expert recommendations. These are intended to serve as references for future work in the development of clinical practice guidelines and expert consensus.

KEY WORDS clinical practice guidelines; expert consensus; evidence level; strength of recommendation; Delphi method

基金项目：上海市卫生健康委员会卫生行业临床研究专项（202240371）；上海申康医院发展中心第二轮促进市级医院临床技能与临床创新三年行动计划——研究型医师创新能力培训项目（SHDC2022CRS053）、第二轮促进市级医院临床技能与临床创新三年行动计划（2023—2025年）CRU 协同数据质量提升项目（SHDC2024CRX032）；上海市皮肤病医院引进人才科研基金项目（2021KYQD01）；上海人才发展基金项目（2021SHRCFZ01）；上海市卫生健康委员会、上海市中医药管理局2025年度中医药标准化项目（2025BZ004）

作者简介：王瑞平，博士，主任医师、教授，研究生导师。研究方向：临床流行病学、生物统计学在临床研究中的应用

通信作者：李斌，博士，主任医师、教授，博士生导师。研究方向：中医药防治皮肤病的基础和临床研究

临床诊疗指南和专家共识在医疗实践中发挥着重要作用，是规范医疗行为、提高诊疗质量的关键工具。为促进新技术、新药物和新疗法的规范应用，国内外各专业学会/协会持续制定和更新临床诊疗指南与专家共识，旨在为临床医师提供科学、合理的诊疗决策依据。然而，在制定过程中，编写团队往往面临如何科学地评估并确定纳入指南或共识的“药物”“技术”和“方法”的循证医学证据等级及专家推荐强度。本文从系统性文献检索出发，深入探讨文献质量评价的具体方法，并据此提出循证医学证据分级和专家推荐强度的标准化评估框架，为临床诊疗指南和专家共识的制定提供参考依据。

1 文献检索策略

以目标疾病治疗药物的中英文“商品名”“通用名”“化学名”和非药物治疗方案的中英文“全称”“简称”“缩略词”等为检索主题词，在包括 PubMed、Embase、The Cochrane Library 和 Web of Science、MEDLINE、中国知网、万方数据和维普资讯等在内的数据库中检索，检索文献周期为“数据库创建的年份至本指南制定的年份”，文献类型包括“药物治疗”和“非药物疗法”的系统综述、荟萃（meta）分析、随机对照试验、队列研究、病例对照研究、病例报道、专家意见等。根据检索到的不同类型的最新文献，认真研读并逐一开展文献质量评价。

2 文献质量评价

对检索到的不同文献类型，采用不同的文献质量评价方法来确定文献质量的高低，具体评价方法如下。

2.1 系统综述和 meta 分析文献质量评价

遵照“PRISMA”声明，采用 AMSTAR2 (a measurement tool to assess systematic reviews 2)量表对“系统综述”和“meta 分析”文献进行质量评价^[1]。AMSTAR2 量表共包含 16 项条目，根据 Beverley J Shea 等量表制作人员的建议，确定 AMSTAR2 量表 7 个关键条目（条目 2、条目 4、条目 7、条目 9、条目 11、条目 13、条目 15）。总体质量评价标准为：①没有或只有一个非关键条目不符合的文献将被评为“高质量”；②有一个以上的非关键条目不符合将被评为“中等质量”；③有一个关键条目存在缺陷将被评为“低质量”；④超过一个关键条目存在缺陷将被评价为“极低质量”。AMSTAR2 量表中的每个条目评价标准如下：如果条目内容报告完全且依据充分，判断为“符合”，计 1 分；

未报告条目相关内容，判断为“不符合”，计 0 分；条目内容回答正确但无充分依据，判断为“部分符合”，计 0.5 分。根据上述评分标准，依据 AMSTAR2 量表评估没有进行 meta 分析的系统评价总分为 13 分，进行 meta 分析的系统评价总分为 16 分（表 1）。

2.2 随机对照临床试验文献质量评价

采用 Jadad 量表评价“随机对照临床试验”文献质量^[2]。评价维度包括“随机”“双盲”和“失访”，目的是反映控制临床试验相应偏倚的能力。Jadad 量表记分原则包括：正确描述随机序列的方法计 2 分，提到 / 随机 / 随机分配 / 随机分组计 1 分，未随机 / 不清楚计 0 分；正确描述双盲计 2 分，提到使用双盲计 1 分，无双盲 / 不清楚计 0 分；报告各组退出与失访病例的原因和例数计 1 分，未充分描述失访计 0 分。如表 2 所示，Jadad 量表满分为 5 分，3 分及以下属于低质量研究，4 分及以上可认为是质量较高的研究。

2.3 病例对照研究和队列研究文献质量评价

采用纽卡斯尔 - 涅太华量表 (Newcastle-Ottawa scale, NOS) 评价病例对照研究和队列研究文献质量^[3]。NOS 量表通过 3 大块共 8 个条目评价病例对照研究和队列研究，具体包括研究人群选择、可比性、暴露评价或结果评价（表 3）。NOS 对文献质量的评价采用星级系统的半量化原则，满分为 9 颗☆，其中得分≥7 颗☆为“高质量研究”；5～6 颗☆为“中等质量研究”；≤4 颗☆为“低质量研究”。

2.4 横断面调查研究文献质量评价

应用美国医疗研究与质量署 (Agency for Healthcare Research and Quality, AHRQ) 量表评估横断面研究的文献质量^[4]。AHRQ 量表包含 11 个条目，采用“是”“否”“不清楚”3 级评价标准，得分与研究质量直接相关（表 4）。AHRQ 量表的每个条目回答“是”得 1 分，“否”或“不清楚”得 0 分；其中条目 5 为反向计分，即“是”或“不清楚”得 0 分，“否”得 1 分。AHRQ 量表总分 11 分，其中得分≥8 分为“高质量研究”，得分 6～7 分为“中等质量研究”，得分≤5 分为“低质量研究”。

3 基于文献质量评价结果形成循证医学证据等级及推荐意见

综合 Cochrane 协作网文献证据等级分级体系、牛津循证医学中心制订的循证医学证据分级和世界卫生组织

• 临床研究规范 •

表 1 AMSTAR2 量表

条目编号	条目内容	符合度判定		
		符合	部分符合	不符合
1	系统评价的研究问题和纳入标准是否基于 PICO 构建	<input type="checkbox"/>		<input type="checkbox"/>
2	制作系统评价前是否制订前期研究方案；制作中若有修订，报告修订的细节	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	研究设计的选择依据是否给予解释	<input type="checkbox"/>		<input type="checkbox"/>
4	是否使用了全面的检索策略	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	研究筛选是否具有可重复性	<input type="checkbox"/>		<input type="checkbox"/>
6	数据提取是否具有可重复性	<input type="checkbox"/>		<input type="checkbox"/>
7	是否提供排除研究的清单以及排除理由	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	是否描述纳入研究的详细基本信息	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	纳入研究的偏倚风险评估方法是否合理	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	是否报告系统评价纳入研究的资金资助信息	<input type="checkbox"/>		<input type="checkbox"/>
11	如果执行 meta 分析，结果合成统计学方法是否合适	<input type="checkbox"/>		<input type="checkbox"/>
12	如果执行 meta 分析，是否评价单个研究偏倚风险对 meta 分析结果的影响	<input type="checkbox"/>		<input type="checkbox"/>
13	在解释和讨论系统评价结果时是否考虑了单个研究的偏倚风险	<input type="checkbox"/>		<input type="checkbox"/>
14	是否对存在的异质性进行满意的解释和讨论	<input type="checkbox"/>		<input type="checkbox"/>
15	如果进行定量合并，是否充分调查了存在发表偏倚的可能性，并讨论发表偏倚对结果的影响	<input type="checkbox"/>		<input type="checkbox"/>
16	是否报告潜在的利益冲突来源，包括目前系统评价获得的资金资助情况	<input type="checkbox"/>	<input type="checkbox"/>	

注：PICO 为患者 / 问题 (patient/problem)、干预 (intervention)、对照 (comparison)、结局 (outcome)。

表 2 Jadad 量表

评价指标	指标描述和解读	计分
随机	未随机 / 不清楚	0
	说明随机	1
双盲	说明随机并描述随机顺序的产生	2
	无双盲 / 不清楚	0
失访	说明双盲	1
	双盲描述充分	2
Jadad 评分	未充分描述	0
	描述	1

GRADE 评价系统^[5]，临床诊疗指南或专家共识的循证医学证据等级如下：1 级（基于同质随机对照试验研究的系统综述或 meta 分析，单个高质量随机对照试验研究，“全或 1”研究证据）；2 级（低质量 RCT、基于队列研究的系统综述）；3 级（病例对照研究、基于病例对照研究的系统综述）；4 级（病例系列报道、横断面研究、低质量病例对照研究和队列研究）；5 级（专家意见、评论或传统临床经验等）。基于上述循证医学证据级别，形成专家推荐意见为：A 级（强推荐）、B 级（推荐）、C 级（可能推荐）、D 级（不推荐，即反对某项措施）（表 5）。

4 推荐意见形成

采用德尔菲专家会商法开展推荐意见的专家共识度评价^[6]，其工作流程见图 1。根据德尔菲专家会商法流程，邀请 15~30 名该研究领域专家参与评价，每位专家对每项推荐意见的同意程度采用 Likert 5 级评分法（1=非常不同意，2=不同意，3=不确定，4=同意，5=非常同意），当评分≥4 分的专家比例≥75% 时，即视为该项推荐意见达成（表 6）。

5 总结

如图 2 所示，在制定临床诊疗指南和专家共识时，编写团队需先从文献检索入手，基于检索到的不同文献类别，分别采用 AMSTAR2 量表对“系统综述”和“meta 分析”文献进行质量评价；采用 Jadad 量表评价“随机对照临床试验”文献质量；采用 NOS 评价病例对照研究和队列研究文献质量；采用 AHRQ 量表评估横断面研究的文献质量。在文献质量评价基础上，综合 Cochrane 协作网文献证据等级分级体系、牛津循证医学中心制订的循证医学证据分级和世界卫生组织 GRADE 评价系统，给出临床诊疗指南和专家共识的循证医学证据等级和专

• 临床研究规范 •

表 3 NOS 量表

病例对照研究			队列研究		
条目	具体条目描述	星级	条目	具体条目描述	星级
人群选择					
1	对于所选择的病例描述是否充分	☆	1	暴露队列的代表性如何	☆
2	该病例是否有代表性	☆	2	非暴露组的选择方法	☆
3	对照组的选择方法	☆	3	暴露的确认方法	☆
4	对照的描述是否充分	☆	4	是否有证实在研究开始时所关注结果确实还没有出现	☆
可比性					
5	以病例组和对照组的设计或分析为基础的可比性	☆☆	5	以设计和分析为基础的队列之间的可比性结果	☆☆
暴露					
6	暴露的确定方法	☆	6	研究对于结果的评价是否充分	☆
7	是否用相同的方法确定对照和病例组的暴露	☆	7	结果发生后随访是否足够长	☆
8	无反应率的相关问题	☆	8	队列群体随访是否充分	☆

表 4 AHRQ 评价量表

条目编号	条目内容	评价选项		
		是	否	不清楚
1	是否明确了资料来源（调查、文献回顾）	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	是否列出暴露组和非暴露组（病例和对照）的纳入及排除标准或参考以往的出版物	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	是否给出纳入患者的时间段	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	如果不是人群来源的话，研究对象是否连续采集	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	评价者主观因素是否掩盖了研究对象的其他方面情况	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	描述了任何为保证质量而进行的评估（如对主观结局指标的检测 / 再检测）	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	解释了排除分析任何患者的理由	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	描述了如何评价或控制混杂因素的措施	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	如果可能，解释了分析中如何处理丢失数据	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	总结了患者的应答率及收据收集的完整性	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	如果有随访，查明预期患者不完整数据所占百分比或随访结果	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

表 5 循证医学证据等级及推荐意见

证据级别	描述	推荐意见
1 级		A 级
1a	基于高质量随机对照试验研究的系统综述或 meta 分析	
1b	单个高质量的随机对照试验研究，非高质量的随机对照试验系统综述或 meta 分析	
1c	“全或无”研究证据，即有该治疗之前，所有患者均死亡，而有该治疗后，有患者能存活。或有该治疗之前，一些患者死亡，而有该治疗之后，无患者死亡	
2 级		B 级
2a	基于高质量队列研究的系统综述或 meta 分析	
2b	单个高质量的队列研究，非高质量的基于队列研究系统综述或 meta 分析，或低质量的随机对照试验研究	
3 级		
3a	基于高质量病例对照研究的系统综述或 meta 分析	
3b	单个高质量的病例对照研究，非高质量的基于病例对照系统综述或 meta 分析	
4 级	病例报道（病例系列研究、横断面调查），低质量的病例对照研究，低质量队列研究	C 级
5 级	未经严格论证的专家意见、评论、传统临床经验	D 级

• 临床研究规范 •

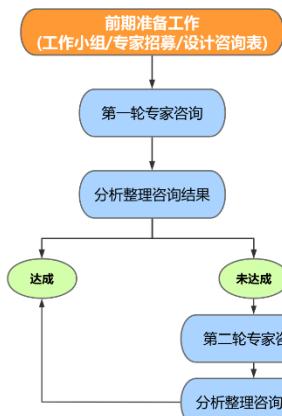


图1 德尔菲专家会商法工作流程图

表6 专家意见 Likert 5 级评分法

问题	证据级别	推荐意见	同意程度评分				
			非常不同意(1)	不同意(2)	不确定(3)	同意(4)	非常同意(5)
1	1a	强推荐	非常不同意(1)	不同意(2)	不确定(3)	同意(4)	非常同意(5)
2	1b	强推荐	非常不同意(1)	不同意(2)	不确定(3)	同意(4)	非常同意(5)
3	2a	推荐	非常不同意(1)	不同意(2)	不确定(3)	同意(4)	非常同意(5)
4	3b	推荐	非常不同意(1)	不同意(2)	不确定(3)	同意(4)	非常同意(5)
5	4	可能推荐	非常不同意(1)	不同意(2)	不确定(3)	同意(4)	非常同意(5)
.....	2b	推荐	非常不同意(1)	不同意(2)	不确定(3)	同意(4)	非常同意(5)

家推荐强度。最后，应用德尔菲专家会商法开展推荐意见的专家共识度评价。

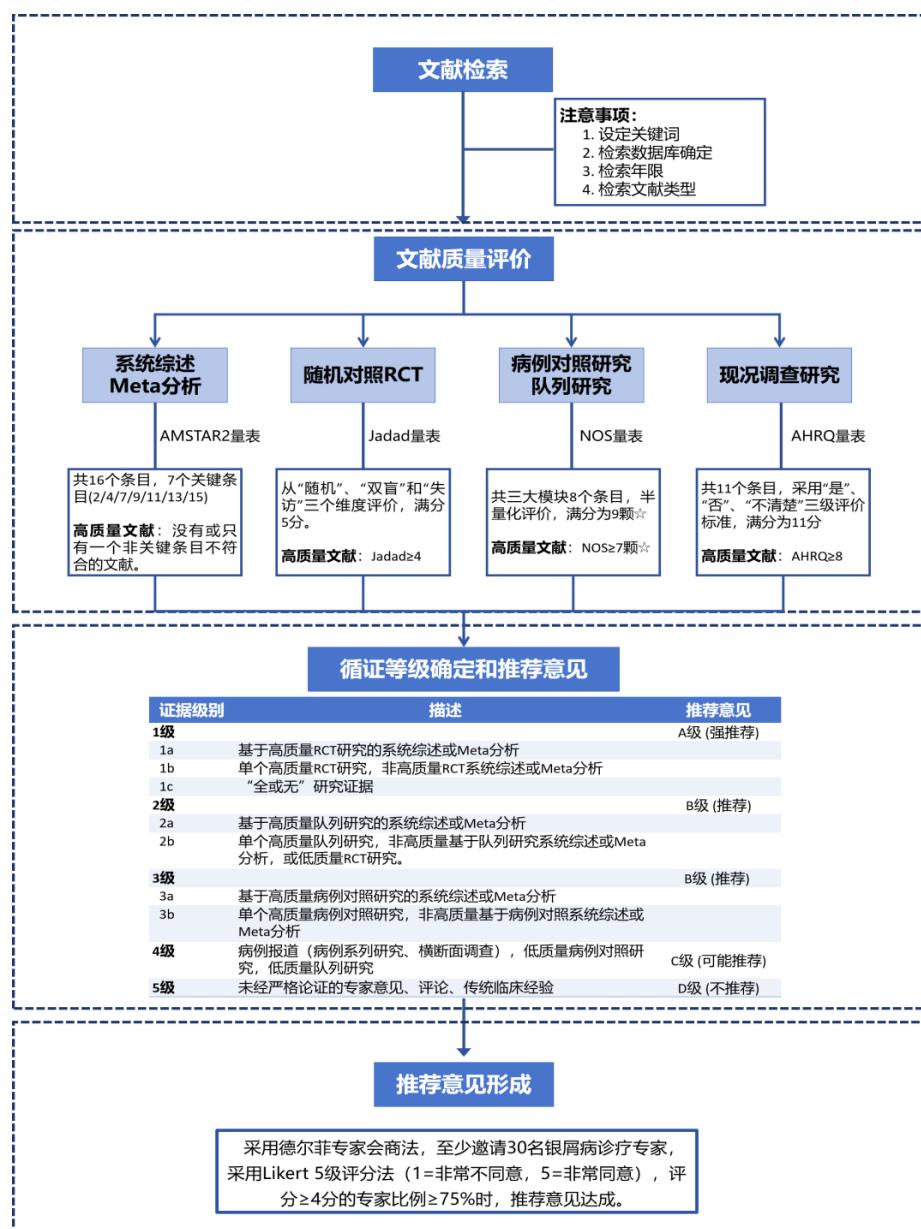


图2 临床诊疗指南和专家共识的制定流程和方法示意图

参考文献

- [1] Leclercq V, Hiligsmann M, Parisi G, et al. Best-worst scaling identified adequate statistical methods and literature search as the most important items of AMSTAR2 (a measurement tool to assess systematic reviews)[J]. J Clin Epidemiol, 2020, 128: 74-82.
- [2] Clark HD, Wells GA, Huöt C, et al. Assessing the quality of randomized trials: reliability of the Jadad scale[J]. Control Clin Trials, 1999, 20(5): 448-452.
- [3] Hartling L, Milne A, Hamm MP, et al. Testing the Newcastle Ottawa scale showed low reliability between individual reviewers[J]. J Clin Epidemiol, 2013, 66(9): 982-993.
- [4] Clancy C. AHRQ: coordinating a quantity of quality[J]. Healthplan, 2003, 44(2): 42-46.
- [5] Zeng LN, Brignardello-Petersen R, Hultcrantz M, et al. GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings[J]. J Clin Epidemiol, 2021, 137: 163-175.
- [6] 高深甚, 王瑞平. 德尔菲专家会商法在定性研究中的应用[J]. 上海医药, 2024, 45(7): 41-44; 70.

(收稿日期: 2025-07-21)

《神经元》: 间歇性禁食可以预激活神经元修复能力, 促进神经受损后修复

周围或中枢神经系统损伤后, 轴突再生受限, 阻碍神经修复。背根神经节 (dorsal root ganglion, DRG) 连接中枢与周围神经, 是研究促轴突再生的理想模型。近日, 一篇发表在《神经元》上的研究显示, 运动、间歇性禁食 (intermittent fasting, IF) 等生活方式因素可在损伤前“预激活”小鼠 DRG 感觉神经元, 增强轴突再生能力。

研究发现, IF 能促进坐骨神经损伤后轴突再生, 并提高血清中喇叭-3-丙酸水平, 增强中性粒细胞向 DRG 的趋化, 从而促进修复。进一步的研究揭示, IF 诱导的代谢变化即 DRG 神经元的瘦素敏感性是神经再生的必要条件。IF 可显著上调瘦素信号, 增强其敏感性; 瘦素过表达可在坐骨神经和脊髓损伤后促进修复。

RNA 测序显示, 在 IF 条件下, DRG 神经元 639 个基因上调、868 个下调。下调通路涉及免疫反应、细胞黏附与趋化; 上调通路集中于脂质代谢、类固醇合成、胆固醇转运及神经发育, 瘦素信号居核心。DRG 神经元表达瘦素受体, 并伴随转录激活因子 3 水平升高, 提示瘦素信号敏化。外源性瘦素同样可促进神经再生, 而删除瘦素受体则该效应消失。

实验表明, 瘦素能促进自由进食小鼠 DRG 神经元轴突生长, 但对 IF 小鼠无进一步作用。体内过表达瘦素, 可在坐骨神经损伤后促进远端神经再生并减轻异常疼痛; 在脊髓损伤中增强轴突生长并减少回缩。转录分析提示, 瘦素过表达可激活依赖环磷酸腺苷的转录网络, 其中腺苷酸环化酶为关键调节因子。

综上, 饮食依赖的神经瘦素信号敏化机制可在不依赖能量代谢的情况下显著促进外周与中枢神经轴突再生。

(摘自奇点糕)

广告索引

封二 上海信谊天平药业有限公司 插二 上海禾丰制药有限公司

封三 上海现代制药股份有限公司 插三 上海上药新亚药业有限公司

封底 颈复康药业集团有限公司 插四 上海复旦复华药业有限公司

插一 上海上药第一生化药业有限公司