**附件5. 医学期刊知识挖掘与服务重点实验室高质量指南共识类语料数据制作规范及管理办法**

1. **编制背景与目的**

规范化和标准化语料加工流程，促进文献资源转化为语料资源，构建形成高质量医学语料数据库，支撑医学人工智能研发与医学语言模型训练，提升医疗场景下人工智能技术应用的效率和效果，并最终推动临床决策支持、智慧科研、精准医疗、健康咨询等场景的智能化升级。

1. **适用范围**

本规范适用于由中华医学会杂志社所属期刊所收录的临床诊疗类指南与共识类文献。在获得版权许可的前提下，其适用范围可扩展至加盟中华医学期刊网的其他期刊。

1. **语料数据制作说明**

### 构建来源

1. 学科方向：以内外妇儿、全科为主，兼顾其他学科方向。
2. 文章类型：指南共识类文献。
3. 来源跨度：来源于同一篇文章。
4. 结构来源：文章的标题、作者、发布时间、关键词、摘要、正文、参考文献等。
5. 内容类型：文本、图片、表格。

### 语料数据类型

1. 客观类问答：指基于文章可以直接找到答案的相关问答对，包含文章中的图表等信息。
2. 主观类问答：指基于文章不能直接找到答案的相关问答对，可包含需要基于文章内容理解归纳总结后生成的问答对，也可包含跨相关文章理解归纳总结后生成的问答对。
3. **语料数据制作标准**

### 客观类问答

1. 专业准确性：回复必须依据最新、权威的临床指南、共识或高质量文献，所述信息无医学错误或误导，术语和概念严格符合专业定义。
2. 证据可溯源性：回答中涉及的数据、结论或推荐应能在相应指南或论文中查证，必要时附引用或指向具体章节/表格，以便验证与追溯。
3. 完整覆盖度：对提出的问题若包含多个子项或多重情景，应逐一回答且不遗漏；对可能存在的变异情形（如特殊人群、合并症等）应酌情补充说明。
4. 可读性与结构性：答案应分段或列举要点，段落层次分明；必要时使用序号或小标题标示逻辑关系；语言流畅，便于专业人士查阅与检索。
5. 实用性与可操作性：建议或结论应具有临床可行性，结合具体操作或决策路径（如诊疗流程、用药方案选择要点等），避免过于笼统或表面化阐述。
6. 一致性与稳定性：同一问题在不同上下文中回答应保持一致；对同一主题的多条语料应避免内容矛盾，保证模型训练后输出稳定不冲突。
7. 时效性与更新机制：依据最新指南或近年重要研究结果；对于快速变化领域（如新药、突发公共卫生事件），需标注信息发布日期并定期更新语料。
8. 安全性与风险提示：如可能存在副作用、禁忌或风险因素，应在回答中明确提示；对潜在严重后果或需进一步专科评估的情况，应建议就医或咨询专家。
9. 中立性与无偏性：回答应避免推广特定厂商、产品或未经验证的疗法；如引用研究或方案，应注明适用范围及局限性，防止商业或地域偏倚影响内容客观性。

### 主观类问答

1. 专业有效性：回答依然需基于权威指南或高质量证据，但对大众健康类或患者教育场景，应兼顾浅显易懂与医学准确，不失严谨性。
2. 同理心与用户导向：对患者或公众提问，应体现同理心，语气友好、尊重受众背景；在建议或指导时注意心理接受度及健康素养水平，提供可行的行动建议。
3. 全面覆盖与情境适配：针对提问时涉及多方面或生活方式因素，应涵盖相关影响因素（如年龄、性别、文化背景、经济条件等），并指出不同情境下的个性化建议或需咨询的专业环节。
4. 清晰易懂与结构化表达：回答应分段或使用小标题，先给出核心结论，再展开细节解释；对于非专业受众，应对专业术语给予简要解释或比喻，便于理解。
5. 实用性与可执行性：建议要具体、可落地，例如生活方式改变方案、监测指标及频次、何时需要专业随访；对长期管理类问题，应提供阶段性目标和评估方法。
6. 证据可追溯与引用透明：即便是科普或主观建议，也应注明其参考来源（如指南章节、权威健康教育材料或高质量研究）；对无法查证或经验性建议，应明确标注“基于临床经验”或“尚缺高质量证据”。
7. 安全性与风险告知：对可能影响健康或存在潜在风险的建议，应加以风险说明和预警；如建议运动或饮食调整，应指出禁忌人群或需谨慎的情形，并建议在专业指导下实施。
8. 文化与伦理敏感性：回答时需顾及不同文化背景或伦理规范（如宗教、饮食习惯），避免冲突或冒犯；对于敏感话题，应给予中立、尊重的回应。
9. 情感支持与激励元素：在适当场景下，提供正向激励或情感支持语言，帮助患者或公众增强信心；同时避免虚假承诺或夸大效果。
10. 可评估性与反馈机制：针对提供的建议，应给出评估进展的指标或方法，便于用户自行或在专业指导下跟踪效果；后续如需调整方案，应提示定期复查或咨询。
11. **指南入选与编制流程**

**1、文献来源与选择标准**

1）来源要求：优先采用已在中华医学期刊网或其他权威平台收录的临床指南、专家共识或技术规范性文件。

2）时效性：指南应发表于近五年内；对于重点领域若标准指南更新周期较长，也可适当考虑稍早但仍具临床代表性和权威性的版本，需在记录中注明发布时间和版本号。

3）权威性评估：核实发布机构（如国家/行业协会、国际学术组织等）及专家组组成，确保内容具备权威背书；避免采纳未经同行评审或小范围发布的不完整文本。尽量选择STAR分级较高的指南类文献。

4）主题覆盖：优选在所需领域具有广泛影响和高引用量的指南；如有多个类似指南，可通过比较内容深度、适用人群、更新频度等维度进行优先级排序。

**2、问答对生成原则**

1）覆盖核心要点：针对每部指南，梳理章节或条目中的关键信息（如适用人群、诊疗流程、推荐级别、禁忌与注意事项、风险提示等），确保问答对涵盖其所有核心内容与临床决策节点。

2）数量与篇幅对应：原则上，1万～5万字的指南生成约100个高质量问答；超过5万字的指南生成约200个问答；对篇幅特别精炼或高度聚焦的小型指南，可酌情减少问答数，但须保证核心要点无遗漏。

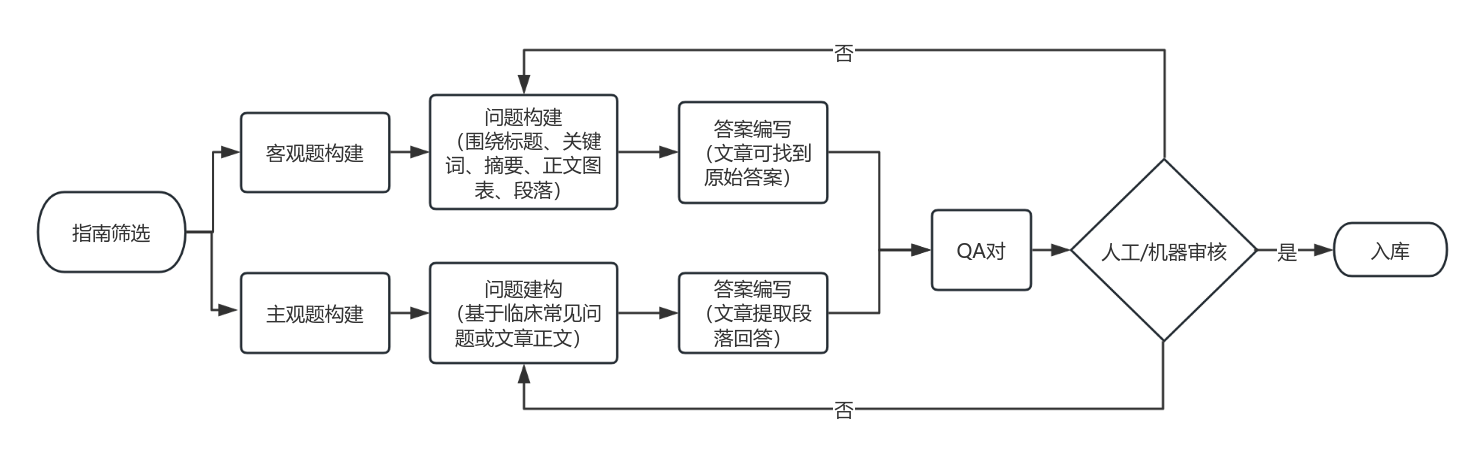
3）问题设计：问题应具备情境或临床场景导向，如“在何种条件下推荐使用X疗法？”、“若患者合并Y疾病，如何调整方案？”等；避免过于宽泛或无实际意义的提问。

4）回答要求：回答须基于指南原文准确提炼，明确说明推荐等级或证据级别；如指南表达含糊，可在回答中注明“指南未明确建议，需结合临床判断或进一步研究”；如涉及数值或阈值，应精准引用并注明出处章节。

5）图表呈现：如指南包含关键信息的图示或表格，且文字难以精准描述，可在语料中直接引用或嵌入相关图表，并注明来源和出处，确保信息完整、准确且可溯源。

**3、编制流程**

指南类语料编制流程如下：



**4、责任主体**

基于指南的语料数据制作须至少设立三名责任人：责任编辑、主笔人和评审专家，且各自职责明确，不得跨角色执行。具体分工如下：

|  |  |
| --- | --- |
| **责任人** | **职责** |
| 责任编辑 | 负责指南初选与立项，邀请主笔人和评审专家，统筹并监督数据制作全流程，确保进度与质量；协调各方沟通，维护项目管理文档。 |
| 主笔人 | 负责将指南内容拆解为符合语料数据要求的问答对，确保内容完整、逻辑清晰，并对文字表达质量负责。 |
| 评审专家 | 对语料数据进行科学性、准确性和完整性审核，提出修改意见；审查文字表述的专业性与合规性，确保数据可靠。 |

1. **数据采集与管理系统**

基于指南的语料数据的提交、评审与入库全程通过中华医学会杂志社一体化学术期刊出版服务平台（MedPress.yiigle.com）执行。执笔人通过平台提交语料数据文档，责任编辑开展初审并邀请外审专家进行同行评审；评审通过后，编辑团队依据规范进行内容加工与格式校验；经终审确认合格后，数据存档入库并纳入后续共享与应用。整个流程在平台中留存完整操作与审核记录，确保可追溯性与合规性。

1. **语料数据制作示例**

### 文献来源

|  |  |
| --- | --- |
| **文献基本信息** | |
| **文题** | 中国血脂管理指南（2023年） |
| **关键词** | 动脉粥样硬化；心血管疾病；血脂；管理 |
| **出处** | 中华心血管病杂志,2023,51(03)：221-255 |
| **全文链接** | https://rs.yiigle.com/cmaid/1449955 |
| **DOI** | 10.3760/cma.j.cn112148-20230119-00038 |
| **通讯单位** | 国家心血管病专家委员会 中华医学会心血管病学分会 中华医学会内分泌学分会 中华医学会糖尿病学分会 中华医学会检验医学分会 中国卒中学会 |

### 客观题

|  |  |
| --- | --- |
| **问题1** | 血脂异常的流行特征是什么？ |
| **回答** | 1. 近几十年来，中国人群的血脂水平、血脂异常患病率明显增加，以高胆固醇血症的增加最为明显。 2. 动脉粥样硬化性心血管疾病超（极）高危人群的降脂治疗率和达标率较低，亟需改善。 |

|  |  |
| --- | --- |
| **问题2** | 我国成人血脂异常患病率的变化趋势是什么？ |
| **回答** | 1. 中国成人血脂异常患病率近年来一直维持在较高水平，且仍然在逐年增长。   2018年全国调查结果显示，≥18岁成人血脂异常总患病率为35.6%，与2015年全国调查的血脂异常患病率相比依然有所上升。   1. 高胆固醇（total cholesterol，TC）血症（TC≥6.2 mmol/L）患病率的增加最明显。   与2015年的数据相比，2018年高TC血症年龄标化患病率增高近1倍（从4.9%增至8.2%）。   1. 高低密度脂蛋白胆固醇（low-density lipoprotein cholesterol，LDL-C）血症患病率持续上升。   2018年≥18岁成人LDL-C≥4.1 mmol/L的比例为8.0%，而2010年和2015年≥18岁成人LDL-C≥4.1 mmol/L的比例分别为5.6% 和7.2%。 |

|  |  |
| --- | --- |
| **问题3** | ASCVD患者降脂治疗率及达标率的现状是什么样？ |
| **回答** | ASCVD人群和患者，血脂异常治疗率及达标率均偏低。  在一级预防的ASCVD高危人群中，降脂药物的治疗率为5.5%。  在已患ASCVD人群中，降脂药物的治疗率为14.5%，LDL-C达标率为6.8%。  此外，在全国246家医院的104516例急性冠脉综合征（acute coronary syndrome，ACS）住院患者中，采用《超高危动脉粥样硬化性心血管疾病患者血脂管理中国专家共识》的标准进行分析显示，75.1%患者为超高危患者，入院时LDL-C达标率（<1.4 mmol/L）为6.6%；在具有出院处方信息的患者中，95.1%的患者出院时仅接受他汀类药物单药治疗。  最新一项9944例包括慢性冠心病、缺血性脑卒中和周围血管疾病的ASCVD患者的随访研究提示，中国ASCVD患者中26%为超高危患者，LDL-C达标率为13%。 |

|  |  |
| --- | --- |
| **问题4** | 他汀类药物不耐受的定义是什么？ |
| **回答** | 1. 指他汀类药物应用后出现与他汀类药物相关的临床不良反应和/或实验室检测指标异常。 2. 同时满足以下4个条件：    1. 临床表现：主观症状和/或客观血液检查不正常；    2. 不能耐受≥2种他汀类药物，其中一种他汀类药物的使用剂量为最小剂量；    3. 存在因果关系；    4. 排除其他原因。 |

### 主观题

|  |  |
| --- | --- |
| **问题5** | 引起血脂升高的原因有哪些？ |
| **回答** | 1. 家族性高胆固醇血症。    1. 成人诊断标准：符合下列三条中的两条：       1. 未经治疗血清低密度脂蛋白胆固醇≥4.7mmol/L。       2. 皮肤黄色瘤或脂性角膜弓（小于45岁）。       3. 一级亲属中有家族性高胆固醇血症或早发冠心病的患者。    2. 儿童诊断标准：       1. 未治疗的血清低密度脂蛋白胆固醇水平≥3.6mmol/L；       2. 且一级亲属中有家族性高胆固醇血症的患者或早发冠心病的患者。 2. 家族性高甘油三酯血症，TG>10 mmol/L。 3. 导致血清脂质和脂蛋白代谢改变的潜在的系统性疾病及代谢状态改变，如肥胖、糖尿病、肾病综合征、甲状腺功能减退症、肾功能衰竭、肝脏疾病、系统性红斑狼疮、糖原累积症、骨髓瘤、脂肪萎缩症、急性卟啉病、多囊卵巢综合征等。 4. 不健康饮食习惯，如摄取富含饱和脂肪酸和胆固醇的饮食可引起胆固醇水平升高，酒精过量可导致高甘油三酯血症。 5. 药物因素，如糖皮质激素、雌激素、维甲酸、环孢素、抗抑郁药物、血管内皮生长因子抑制剂、芳香化酶抑制剂等。 |

|  |  |
| --- | --- |
| **问题6** | 如何通过调整生活方式来改善血脂异常呢？ |
| **回答** | 1. 降脂治疗中首先推荐健康生活方式，包括合理膳食、适度增加身体活动、控制体重、戒烟和限制饮酒等。 2. 合理膳食对血脂影响较大。    1. 即限制饱和脂肪酸及反式脂肪的摄入，增加水果，蔬菜、全谷薯类、膳食纤维及鱼类的摄入。 3. 国外指南推荐的健康膳食模式主要是DASH（dietary approaches to stop hypertension）膳食（美国）和地中海膳食（欧洲）。 4. 中国学者提出了中国心脏健康膳食模式，显著降低血压。 5. 对ASCVD中高危人群和高胆固醇血症患者每天膳食胆固醇摄入量应在300 mg以下。 |

|  |  |
| --- | --- |
| **问题7** | 怎样合理应用降脂药物进行血脂达标管理？ |
| **回答** | 1. 当生活方式干预不能达到降脂目标时，应考虑加用降脂药物。 2. 他汀类药物是降胆固醇治疗的基础。    1. 其剂量增倍，LDL-C降低效果只增加6%；    2. 有潜在的副作用，如肝功能损害、肌病及新发糖尿病等。 3. 我国人群对大剂量他汀类药物的耐受性较欧美人群差，推荐起始使用常规剂量或中等强度他汀类药物。 4. 对于他汀类药物不耐受者可使用天然降脂药血脂康作为起始降脂治疗。 5. 当他汀类药物或血脂康不能使LDL-C达标时，可联合使用非他汀类降脂药物，如胆固醇吸收抑制剂或PCSK9抑制剂。    1. 最近亚洲人群研究显示，在ASCVD患者中，与使用高强度他汀类药物比较，中等强度他汀类药物联合依折麦布有更高的LDL-C达标率和更好的耐受性，且ASCVD事件有降低趋势。    2. 中等强度他汀类药物联合非他汀类药物可替代高强度他汀类药物，疗效和安全性更好。 6. 对于超高危患者，当基线LDL-C较高（未使用他汀类药物患者，LDL-C≥4.9 mmol/L；或服用他汀类药物患者，LDL-C≥2.6 mmol/L），预计他汀类药物联合胆固醇吸收抑制剂不能使LDL-C达标时，可考虑直接采用他汀类药物联合PCSK9抑制剂，以保证患者LDL-C早期快速达标。 7. 提早使用PCSK9单抗可更早和更显著降低ASCVD风险，且长时间使用（≥7年）具有良好的安全性。 |

|  |  |
| --- | --- |
| **问题8** | 治疗过程中如何知道是否达标？ |
| **回答** | 1. 降脂治疗中监测的目的：    1. 观察是否达到降脂目标值；    2. 了解药物的潜在不良反应。 2. 监测的频率：    1. 采取饮食控制等非药物治疗者，开始的3~6个月应复查血脂水平；如血脂控制达到建议目标值，则继续非药物治疗；仍需每6个月至1年复查1次；长期达标者可每年复查 1次。    2. 首次服用降脂药物者，应在用药4~6周内复查血脂、肝酶和肌酸激酶（creatine kinase，CK）。如血脂参数能达到目标值，且无药物不良反应，逐步改为每3~6个月复查 1次。如治疗1~3个月后，血脂仍未达到目标值，需及时调整降脂药物剂量或种类或联合应用不同作用机制的降脂药物。 3. 每当调整降脂药物种类或剂量时，都应在治疗 4~6周内复查。 |

1. **编制单位及修订时间**

编制单位：中华医学会杂志社 国家新闻出版署医学期刊知识挖掘与服务实验室。

本规范编制日期：2025年6月15日。

最新修订日期：2025年6月15日。