

面向医学文本处理的医学实体标注规范

张欢，宗源

指导教师：常宝宝、穗志方、昝红英、张坤丽

{zhang_huan,zongyuan,chbb,szf}@pku.edu.cn

{iehyzan,ieklzhang}@zzu.edu.cn

目录

一、引言.....	3
二、相关工作.....	4
三、总体原则.....	6
3.1 简单性原则.....	6
3.2 易操作性原则.....	6
3.3 一致性原则.....	7
四、医学实体体系.....	8
五、医学实体标注细则.....	10
5.1 疾病.....	10
5.1.1 疾病或综合征.....	10
5.1.2 受伤或中毒.....	10
5.1.3 器官或细胞损伤.....	11
5.2 临床表现.....	13
5.2.1 症状.....	13
5.2.2 体征.....	14
5.3 医疗程序.....	21
5.3.1 检查程序.....	21
5.3.2 治疗或预防程序.....	21
5.4 医疗设备.....	23
5.4.1 检查设备.....	23
5.4.2 治疗设备.....	24
5.5 药物.....	25
5.6 医学检验项目.....	27
5.7 身体.....	29
5.7.1 身体部位.....	29
5.7.2 身体物质.....	30
5.8 科室.....	32
5.9 微生物类.....	32
六、实体分类混淆处理.....	33
6.1 疾病(dis)和临床表现(sym).....	33
6.2 医疗程序(pro)和医疗设备(equ).....	34
6.3 药物(dru)和身体(bod).....	35
6.4 医学检验项目(ite)和医疗程序(pro).....	36
6.5 医学检验项目(ite)和身体(bod).....	37
七、结束语.....	38
参考文献.....	39

一、引言

近年来，互联网和数字化已在众多行业带来颠覆性变革，医疗健康领域也不例外。提起医疗，可能映入脑海的关键词就是看病难、看病贵，这和长期以来医疗行业供需不匹配、信息不平衡等因素有很大关系。而智慧医疗的到来，在很大程度上改进医院的管理及运营模式、改进对大众的医疗服务。

对于医学领域的大量自然语言文献，例如医学教材、医学百科、临床路径、病历、医学期刊、检验报告等，这些医学文本其中蕴含了大量的专业知识和丰富的有用信息，比如医学术语等。如何将海量的有用信息充分利用是目前医学领域的一大难题，并且仅依靠人力去识别和提取信息必将是一项耗时又耗力的工程。如果利用人工智能的技术，让机器“读懂”医学数据，必然可以显著提高临床科研效率和质量。在智慧医疗的普及过程中，利用“机器读懂患者问题进行自动回答”的方式，不仅可以大幅减少患者服务所需的人力，同时也将之前只有大城市拥有的医疗资源分配到更多更广泛的区域。要想让计算机“读懂”医学数据，核心在于让计算机能在众多医学文本中提取出关键信息，这就涉及到了自然语言处理技术，主要包括命名实体识别、关系提取等。医学领域中的命名实体识别指的是将重要的医学实体，如疾病、症状等从医学文本中抽取出来，这个步骤也是医学关系提取等各项任务的基础。

命名实体识别的主要技术方法分为：基于规则和词典的方法、基于统计的方法、二者混合的方法。基于规则和词典的方法是命名实体识别中最早使用的方法，但规则往往过于依赖知识库，故而充满局限性。基于统计的方法利用人工标注的语料进行训练，标注语料时不需要广博的语言学知识，且可以在较短时间内完成。基于统计的方法现已成为目前研究的主流方法。对于医学命名实体而言，国内没有统一的医学命名实体标注规范，医学共享语料库人处于空白状态，这对医学领域各项 NLP 任务的进展造成了巨大阻力。目前针对不同的标注任务，其医学实体标注规范却是各有不同，医学实体的分类也是大不相同，这也导致缺乏类似通用场景的大规模规范的标注数据。因此，现在最难的是标注的标准问题，对于很

多词的实体类别判断很难达成共识,如何界定不同实体间的涵盖范围成为首要解决之任务。

本文正是基于这样的前提,提出统一化的面向医学文本处理的医学实体标注规范,将医学实体划分为九大类,包含“疾病”、“临床表现”、“医疗程序”、“医疗设备”、“药物”、“医学检验项目”、“身体”、“科室”和“微生物类”。本文分为6个部分:第二部分总结医学命名实体标注的背景;第三部分说明本文建立的医学命名实体体系;第四部分详细介绍了本文对医学命名实体的分类及界定;第五部分阐述对不同的医学命名实体间的混淆处理;第六部分是本文的结束语。

二、相关工作

一体化医学语言系统(Unified Medical Language System; UMLS)是美国国立医学图书馆(NLM)自1986年起研究和开发的一体化医学语言系统^[1],是对生物医学科学领域内术语词表的统一汇编,并提供了UMLS数据库,如超级叙词表、语义网络和专家词典,以及相关软件工具,如MetamorphoSys、MMTx等。SNOMED CT(Systematized Nomenclature of Medicine -- Clinical Terms)医学系统命名法——临床术语,是当前国际上广为使用的一种临床医学术语标准,提供了一套全面统一的医学术语系统,涵盖大多数方面的临床信息,如疾病、所见、操作、微生物、药物等。2010 i2b2/VA challenge会议发布了电子病历命名实体的分类^[2],该会议参考UMLS定义的语义类型,将医学实体分为3种:医疗问题(Medical Problem)、检查(Test)和治疗(Treatment)。Roberts^[3]等人随即选择了50份临床记录、X射线和病理报告进行标注,将这些医学文本中的医学实体分为6种:状况(Condition)、药物(Drug)、干预(Intervention)、部位(Locus)、检查(Investigation)、结果(Result)。South^[4]使用了316例炎症性肠病的临床记录进行标注,其中医学实体种类分为4种:体征或症状(Signs or symptoms)、诊断(Diagnoses)、程序(Procedures)和药物(Medications)。

相比国外对医学语料库的构建以及相关任务的展开,国内没有公开可获得的面向医学命名实体识别的数据集。2014年Lei^[5]等人使用北京协和医院2013年的

电子病历进行标注,其中医学实体分为4种:医疗问题、治疗程序、药物和检查。2014年,Xu^[6]使用一家中国医院提供的336个出院总结进行标注,其中医学实体同样分为4种:医疗问题、治疗程序、药物和检查。2016年,哈尔滨工业大学团队^[7]使用来自哈尔滨医科大学附属第二医院的122个科室的电子病历进行标注,并且制定了新的中文电子病历命名实体和实体关系标注规范。该规范将医学实体分为5种:疾病、疾病诊断分类、症状、检查和治疗。2019年,Gao^[8]等人使用了255份来自中国湖南省某著名医院的真实入院记录进行标注,并在其论文中提出的新的标注方案,将医学实体分为9种:医疗发现、症状、时间词、疾病、身体部位、药物、治疗、实验室检查和(非实验室)检查。

目前为止,人们对医学文本的研究主要集中在电子病历、入院记录、出院总结、病理报告等,并在其任务领域内制定各自的医学实体标注规范,故而不存在统一的医学命名实体标注规范。针对这个问题,本文旨在提出统一化的面向医学文本处理的医学实体标注规范,并将医学实体划分为九大类,包含“疾病”、“临床表现”、“医疗程序”、“医疗设备”、“药物”、“医学检验项目”、“身体”、“科室”和“微生物类”。回顾以上中英文医学文本语料库标注准则,大部分都将“检查”和“治疗”视为医学实体。本文认为,“检查”和“治疗”应该是作为医学实体之间的某种关系,比如“某种药物治疗某种疾病”、“使用某种设备检查身体部位”或“检查哪种检验项目”等等,从而建立“底层”独立实体之间的“高阶”关系,“实体”和“关系”是相互独立的,这样的优势在于让标注者能够更好地理解实体的概念。本文引入的“医疗程序”和“医疗设备”实体就是将“检查”和“治疗”更加“实体化”和“概念化”。一种医疗程序既可以作为治疗某种疾病的过程,也可以作为检查某种疾病是否存在的过程;一种医疗设备同样可以用来治疗疾病,也可以用来检查疾病或身体;等等。此时,疾病、医疗程序、医疗设备和身体作为“底层”实体,而检查和治疗作为“高阶”关系,这样使标注者能够真正建立起医学的概念,仅仅是研究实体本身的含义,而不关注实体之间的关系。这种实体的细致划分一方面是为了概念细粒度化,另一方面为后续的实体关系提取提供了良好的数据基础。另外,“医学检验项目”、“身体”、“科室”和“微生物类”同样作为医学领域中的重要实体,在后续的医学文本处理以及应用中起着至关重要的作用,故而这些实体同样应该标注。

三、总体原则

3.1 简单性原则

本文先将医学实体划分为九大类，包含“疾病”、“临床表现”、“医疗程序”、“医疗设备”、“药物”、“医学检验项目”、“身体”、“科室”和“微生物类”，并详细介绍了各个医学命名实体的涵盖的子类，阐述实体间的混淆处理，用大量示例举例说明。标注者无需太多专业知识，实体类别定义简易明了，方便标注者理解和区分。

3.2 易操作性原则

对于医学文本的命名实体识别，标注者需严格遵从本文提出的医学实体标注规范，使用“[named-entity]tag”的方式进行紧密标记（左右括号与标记实体首末字符之间无空格，括号需成对出现），若出现标注实体的英文缩写、中文简称或者俗称，均需要标注，形式标记方案如表 3-1。鉴于“临床表现”实体内部的复杂性，本文规定该实体内部允许嵌套标注，即若“临床表现”实体内部存在其他 8 种实体，标注者也应该将其标注出来。此外，对于除“临床表现”之外的医学实体，标注应当遵循“最大单位标注法”，即若一个实体内包含其他实体，则标注“最大”的实体，不做嵌套标注。

表 3-1 形式标记方案

范畴	医学实体一级分类	备注	一级标签	医学实体二级分类	二级标签
医学	疾病	disease	dis	疾病名或综合征名	dis dis-ds
				受伤或中毒	dis dis-ip
				器官或细胞损伤	dis dis-ch
	临床表现	symptom	sym	症状	sym sym-sd
				体征	sym sym-sn
	医疗程序	procedure	pro	检查程序	pro pro-cp
				治疗或预防程序	pro pro-tp
	医疗设备	equipment	equ	检查设备	equ equ-ce
				治疗设备	equ equ-te
	药物	drug	dru	-	-
医学检验项目	item	ite	-	-	
身体	body	bod	身体部位	bod bod-pt	

				身体物质	bod bod-mt
	科室	department	dep	-	-
	微生物类	microbes	mic	-	-

考虑到人工标注的工作量和复杂性，本规范提供的标注语料仅包含医学实体一级分类信息，也就是说标注语料仅涵盖“疾病”、“临床表现”、“医疗程序”、“医疗设备”、“药物”、“医学检验项目”、“身体”、“科室”和“微生物类”九种医学实体。为提供更好的理解和展示，“疾病”、“临床表现”、“医疗程序”、“医疗设备”、“身体”包含的子类以二级标签的形式在此规范的举例中体现。

3.3 一致性原则

医学实体的标注包括实体类型和实体边界两个部分。医学实体的分词存在很多歧义，如何切分较长的疾病或药物名称等，这给标注工作带来了很大困难。对于除“临床表现”这个复杂的医学实体外，人们往往关注的是医学实体的含义，比如什么疾病、什么药物、哪个科室等等，这类医学实体内部无需分词，仅作为一个整体来看待。因此，本文主张“医学文本总体遵从《北京大学现代汉语语料库基本加工规范（2002版）》^[9]、除临床表现之外的实体类别中不存在分词”，这将大大提高标注的效率和质量。另外，鉴于为保证实体意义的完整性和可理解性，所有的实体可以是一个词、短语和句子，实体中可包含标点符号。也就是说，当标点符号存在某种意义时，标注者同样应该将其标注。

四、医学实体分类体系

首先，本文将医学实体划分为九大类，包含“疾病”、“临床表现”、“医疗程序”、“医疗设备”、“药物”、“医学检验项目”、“身体”、“科室”和“微生物类”。其次，对于“疾病”、“临床表现”、“医疗程序”、“医疗设备”、“身体”此类医学实体如图 4-1 所示。故本文提出的规范对于医学实体的划分上更加细致，这也有便于未来医学实体关系提取等各项工作的开展。本文借鉴 UMLS 语义类型界定实体涵盖的范围，但不局限于 UMLS 的定义。（相关工作 得说）

第一类实体是疾病，疾病是指导致病人处于非健康状态的原因或者医生对病人做出的诊断，并且是能够被治疗的^[10]。包括疾病或综合征、中毒或受伤、器官或细胞受损，其对应的 UMLS 语义类型有疾病或者综合征（disease or syndrome）、中毒或受伤（injury or poisoning）等；

第二类实体是临床表现，临床表现是疾病的表现，泛指患者不适感觉以及通过检查得知的异常表现。主要包括症状、体征，其对应的 UMLS 语义类型有症状或体征（sign or symptom）、异常检查结果（abnormal test results）等；

第三类实体是医疗程序，在本文中，医疗程序泛指为诊断或治疗所采取的措施、方法及过程。主要包括检查程序、治疗或预防程序，其对应的 UMLS 语义类型有化验过程（laboratory procedure）、治疗或预防过程（therapeutic or preventive procedure）、等；

第四类实体是医疗设备，在本文中，医疗设备泛指为诊断或治疗所使用的工具、器具、仪器等。主要包括检查设备、治疗设备，其对应的 UMLS 语义类型有医疗设备（medical device）、药物传输设备（drug delivery device）等；

第五类实体是药物，药物是指用来预防、治疗及诊断疾病的物质，其对应的 UMLS 语义类型有临床药物（clinical drug）、抗生素（antibiotic）等；

第六类实体是医学检验项目，医学检验项目是指检查涉及到的体液检查项目、重要生理指标以及其他检查项目，本文规定“医疗检验项目”主要针对人体而言，是能够通过设备或实验检测出的项目，并且是能够被量化，有其对应的测量值或指标值。其对应的 UMLS 语义类型有实验室检查（laboratory test）等；

第七类实体是身体，身体泛指细胞、组织、及位于人体特定区域的由细小物质成分组合而成的结构、器官、系统、肢体，另外包括身体产生或解剖身体产生的物质等。主要包括身体部位、身体物质，其对应的 UMLS 语义类型有身体部位 (body part)、组织 (organ)、组织成分 (organ component) 等；

第八类实体是科室，科室主要是指医院或医疗机构所设有的科室其对应的 UMLS 语义类型有医疗保健相关组织 (health care related organization) 等；

第九类实体是微生物类，微生物类包括细菌、病毒、真菌以及一些小型的原生生物、显微藻类等在内的一大类生物群体，另外包括微生物类产生的毒素、激素、酶等，其对应的 UMLS 语义类型有细菌 (virus)、真菌 (fungus) 等；

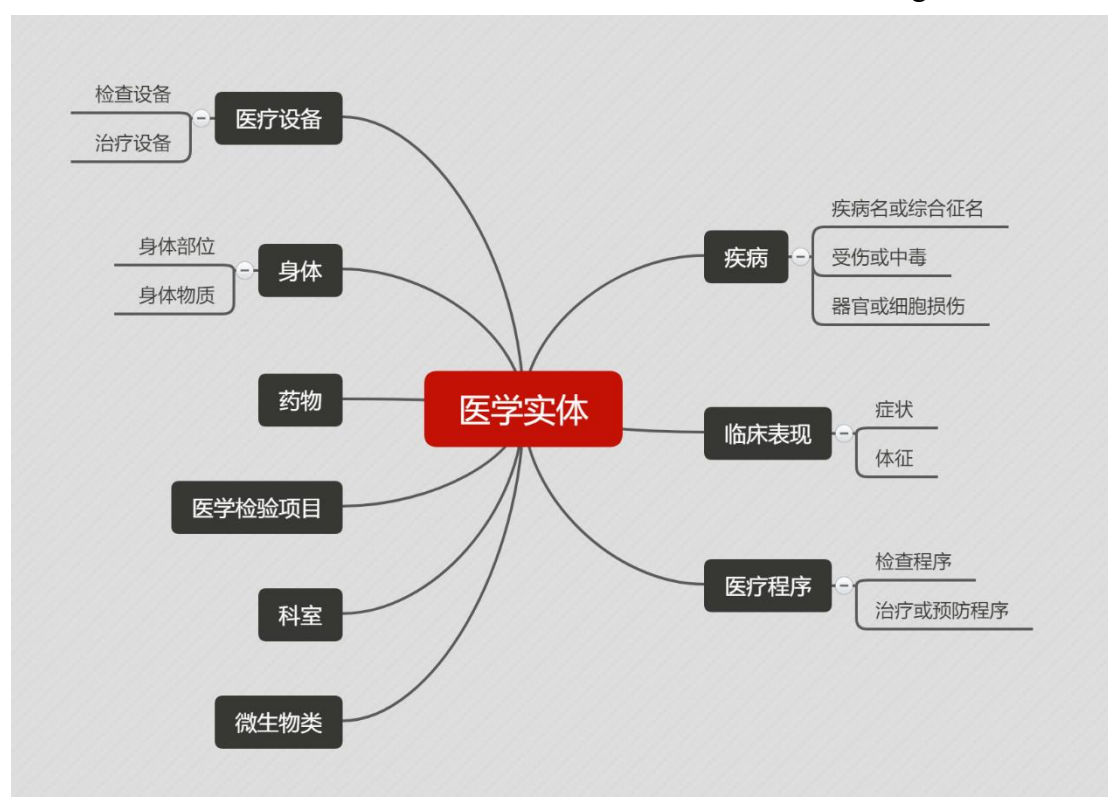


图 4-1 医学命名实体架构图

五、医学实体标注细则

5.1 疾病

“疾病”是指导致病人处于非健康状态的原因或者医生对病人做出的诊断，并且是能够被治疗的。“疾病”（tag为dis）实体分为疾病名或综合征名（tag为dis|dis-ds）、受伤或中毒（tag为dis|dis-ip）、器官或细胞损伤（tag为dis|dis-ch）三个子类。

5.1.1 疾病名或综合征名

疾病或综合征是指疾病或综合征名称。比如高血压、肺炎、心脏病、败血症等。此类医学实体的二级分类标签为dis|dis-ds。

- ① 多数 病儿 在 儿童期 死于 [肺炎]dis|dis-ds 或 [心脏病]dis|dis-ds ， 少数 类型 可 存活 至 成人 。
- ② 晚期 并发症 包括 [糖尿病]dis|dis-ds 、 [动脉血栓]dis|dis-ds 、 [胰腺癌]dis|dis-ds 和 [腹部]bod|bod-pt 其他 [肿瘤]dis|dis-ds 。
- ③ 常见于 [局灶节段型肾小球肾炎]dis|dis-ds 、 [IgA 肾病]dis|dis-ds 、 [系膜增殖性肾炎]dis|dis-ds 、 [草酸盐沉积症]dis|dis-ds 以及 [溶血尿毒综合征]dis|dis-ds
- ④ 测定 结果 的 分析 ： [肺炎]dis|dis-ds 、 [败血症]dis|dis-ds 、 严重 [先天性心脏病]dis|dis-ds 或 [畸形]dis|dis-ds 影响 新生儿 代谢 和 循环 功能

5.1.2 受伤或中毒

患者在受伤或中毒后，对人体造成某种伤害，导致患者处于非健康状态。此类医学实体的二级分类标签为dis|dis-ip。

- ① 患者 经 初步 处理 后 应 立即 送 医院 住院 作 进一步 检查 治疗 。
- ② [高血钾型肾小管酸中毒]dis|dis-ip 又 称 [全远端型肾小管酸中毒]dis|dis-ip ， 是 由于 [[醛固酮]bod|bod-mt 分泌 不足]sym|sym-sn 或

[[肾小管]bod|bod-pt 对其反应低下]sym|sym-sn 所致 [高血氯性代谢性酸中毒]dis|dis-ip 及 [持续高钾血症]dis|dis-ds 。

5.1.3 器官或细胞损伤

器官、细胞等发生异常或损伤后，如果能够危及人的机体，此时虽然它们属于身体的一部分，但是已成为一种致病因素，危害人体健康。此类医学实体的二级分类标签为 dis|dis-ch。

- ① 是否有 [颈部受伤]dis|dis-ip 和其他 [脏器损伤]dis|dis-ch
- ② 小儿时期 [原发性癫痫]dis|dis-ds 最为多见（其表现详见本篇第五章）。其他还有 [颅内出血]dis|dis-ch（[产伤]dis|dis-ip、[窒息]sym|sym-sd、[外伤]dis|dis-ip，[颅脑损伤]dis|dis-ch（[外伤]dis|dis-ip 史）

注意：

- (1) “疾病”实体内部不存在分词，可包含标点符号。
- (2) 与疾病类实体有关的特殊尾词，主要包括“病”、“症”，“癌”，“炎”，在标注时应该与此类尾词一起标注。
 - ① 常见病因有 [先天性心脏病]dis|dis-ds、[暴发性心肌炎]dis|dis-ds、严重 [心律失常]dis|dis-ds、[心脏压塞]dis|dis-ds 和 [急性肺梗死]dis|dis-ds 等。
 - ② [上消化道出血]dis|dis-ch 常见原因：[溃疡病]dis|dis-ds、[炎症]dis|dis-ds、[胃底食管静脉曲张]dis|dis-ds、[反流性食管炎]dis|dis-ds、[Mallory-Weiss 综合征]dis|dis-ds、[胆道出血]dis|dis-ch 及 [胰腺炎]dis|dis-ds。[下消化道]bod|bod-mt：[细菌性肠炎]dis|dis-ds、[炎症性肠道疾病]dis|dis-ds、[息肉]dis|dis-ds 及 [痔]dis|dis-ds。
- (3) 一般有些疾病名称很长，前面会有“XX 性”、“XX 状”、“XX 型”等，以及身体部位（一个或多个）的修饰，在保证疾病具体性的情况下，在标注时应该与这些前缀一起标注。
 - ① 对某些 [病毒性肠炎]dis|dis-ds 有诊断价值，如 [轮状病毒性肠

炎]dis|dis-ds 和 [诺沃克病毒性肠炎]dis|dis-ds 等。

② 有些引起明确的相关性疾病，如 [HAV 感染]dis|dis-ds 与 [急性黄疸型肝炎]dis|dis-ds；[轮状病毒感染]dis|dis-ds 与 [季节性婴幼儿腹泻]dis|dis-ds

③ [会阴和横纹肌肉瘤]dis|dis-ds 少见，通常为腺泡型，对 [放疗]pro|pro-tp 和 [化疗]pro|pro-tp 反应极差，存活率大约为 20%。（这里，“会阴”和“横纹肌肉”都是身体部位，而“瘤”表示这两处身体部位发生的疾病，故应该将两个部位与“瘤”整体标注为“疾病”）

④ 晚期并发症包括 [糖尿病]dis|dis-ds、[动脉血栓]dis|dis-ds、[胰腺癌]dis|dis-ds 和 [腹部]bod|bod-mt 其他 [肿瘤]dis|dis-ds（虽然“肿瘤”前面有“腹部”修饰，但是具有不具体性，故应该分开标注）

⑤ 小儿患有 [肝脏、肾脏、甲状腺疾病]dis|dis-ds（“肝脏、肾脏、甲状腺疾病”有多个身体部位作为前缀，应该整体标注为“疾病”）

⑥ [皮肤、五官、消化道、及呼吸道黏膜损伤]dis|dis-ch

(4) 大部分统称均标注，比如营养性疾病、代谢性疾病、化脓性和非化脓性综合征等，此类虽然是统称，但是有对应的疾病范畴，所以应该标注。特殊情况：像常见病、多发病、疾病等单独出现时，此类统称范围太广，不应标注。

① [高血压]dis|dis-ds 是严重危害人类健康的常见病、多发病，是引起成人 [心脑血管疾病]dis|dis-ds 的主要危险因素之一。（“高血压、心脑血管疾病”标注为“疾病”；“常见病、多发病”不标注）

② 将缺铁性贫血列为全球四大 [营养性疾病]dis|dis-ds 之一。

(5) 一种疾病有若干种分型时，“疾病+分型”或“分型+疾病”整体标注，否则不标注。

① 可将 [MPS]dis|dis-ds 分为 I~VII 型，除 [MPSII 型]dis|dis-ds 为 X 连锁隐性遗传外（“疾病+分型”整体标注，“分型”单独出现不标注）

- ② [会阴和横纹肌肉瘤]dis|dis-ds 少见，通常为腺泡型，对放疗和化疗反应极差，存活率大约为 20%。（“腺泡型”作为疾病分型单独出现，则不标注）

5.2 临床表现

临床表现是疾病的表现，泛指患者不适感觉以及通过检查得知的异常表现。这两种都是疾病的临床表现，但是又明显不同。“临床表现”（tag为sym）实体分为症状（tag为sym|sym-sd）和体征（tag为sym|sym-sn）三个子类。

5.2.1 症状

症状是指病人自己向医生陈述（或是别人代述）的不适或痛苦表现。通常是病人主观感觉的不适，如腹痛、头晕等，或是自己发现的病理改变，如血尿便血、活动障碍等。此类医学实体的二级分类标签为 sym|sym-sd。

- ① 伴随 症状 包括 [恶心]sym|sym-sd 、 [腹痛]sym|sym-sd 、 [头痛]sym|sym-sd 、 [畏光]sym|sym-sd 及 [倦怠]sym|sym-sd ；
- ② [关节]bod|bod-pt 症状 多 表现 为 [肿胀]sym|sym-sd 、 [疼痛]sym|sym-sd 、 [发热]sym|sym-sd 、 [触痛]sym|sym-sd 及 [活动 障碍]sym|sym-sd 。
- ③ [异体血制品]dru 输入 时 可 出现 [过敏 反应]sym|sym-sd （ [[输血]pro|pro-cp 反应]sym|sym-sd ） ， 出现 [发热]sym|sym-sd 、 [皮疹]sym|sym-sd 、 [瘙痒]sym|sym-sd 、 [哮喘]sym|sym-sd 等 反应 ， 严重 时 出现 [过敏性 休克]sym|sym-sd 。
- ④ 有 相当 多 的 病人 表现 为 [胃纳 减退]sym|sym-sd 、 [恶心]sym|sym-sd 、 [呕吐]sym|sym-sd 、 [腹痛]sym|sym-sd 、 [[关节肌肉]bod|bod-mt 疼痛]sym|sym-sd 、 [呼吸 深快]sym|sym-sd 、 [呼气 中 带有 酮味]sym|sym-sd ， [神志 萎靡]sym|sym-sd 、 [嗜睡]sym|sym-sd 、 [反应 迟钝]sym|sym-sd ， 严重 者 可 出现 [昏迷]sym|sym-sd 。
- ⑤ 轻型 患者 一般 [休克]sym|sym-sd 甚至 [死亡]sym|sym-sd 。

5.2.2 体征

体征是指医生观察到的或者通过检查程序或设备检查到的发生于病人的异常变化以及异常检查结果。通常是指医师利用自己的感官（视触叩听）或者医疗器具（血压计叩诊锤等医疗设备）发现的病人的病理生理变化。此类医学实体的二级分类标签为 sym|sym-sn。

- ① 随着神经影像学的发展，临床发现有些 [头颅]bod|bod-pt 较大的婴儿，行 [头颅 CT]pro|pro-cp 和 [MRI 检查]pro|pro-cp 显示 [[双侧额部]bod|bod-pt 或 [额顶部]bod|bod-mt 有 [蛛网膜下腔]bod|bod-mt 增宽]sym|sym-sn，没有或仅有 [轻度 [脑室]bod|bod-pt 扩大]sym|sym-sn
- ② [颅脑 CT]pro|pro-cp 平扫可见 [团块状混杂密度病灶，边缘不清，可有钙化影及局部脑萎缩]sym|sym-sn；增强扫描可见 [异常 [血管]bod|bod-mt 强化影]sym|sym-sn。
- ③ 但 [X线胸片]pro|pro-cp 上已经可以出现 [[两肺]bod|bod-pt 细颗粒影]sym|sym-sn，为 [[间质]bod|bod-mt 水肿]sym|sym-sn 的征象。
- ④ 有 [大型缺损、分流量大、[左心室]bod|bod-pt 超容]sym|sym-sn 的小儿，[胸片]pro|pro-cp 表现为 [心影向左下扩大]sym|sym-sn、[[左心房]bod|bod-pt 扩大]sym|sym-sn、[肺野淤血]sym|sym-sn；如出现 [[肺动脉]bod|bod-pt 高压]sym|sym-sn，[[肺动脉干]bod|bod-pt 突出]sym|sym-sn，[[右室]bod|bod-pt 肥厚]sym|sym-sn，[[心尖]bod|bod-pt 上翘]sym|sym-sn。

注意：

- (1) “临床表现”实体内部允许分词，可包含标点符号。
- (2) 在临床表现实体内部，若包含除“临床表现”之外的其他实体（“疾病”、“身体”、“医学检验项目”等），在允许分词的情况下，内部实体应作为方括号嵌套成分（如“[[肺动脉干]bod|bod-pt 突出]sym”）；当身体部位为单字时，此类情况取决于身体部位是否是词的一部分，例如：背痛、头痛中的“背”和“头”不用单独标出，因为他们是词的一个部分。因此，只要身体部位能够与对应的修饰成分组成词，则不标注身体部位。

- ① [胸片]pro|pro-cp 表现为 [心影向左下扩大]sym|sym-sn 、 [[左心房]bod|bod-pt 扩大]sym|sym-sn 、 [肺野淤血]sym|sym-sn ；如出现 [[肺动脉]bod|bod-pt 高压]sym|sym-sn ， [[肺动脉干]bod|bod-pt 突出]sym|sym-sn ， [[右室]bod|bod-pt 肥厚]sym|sym-sn ， [[心尖]bod|bod-pt 上翘]sym|sym-sn 。
- ② 行 [头颅 CT]pro|pro-cp 和 [MRI 检查]pro|pro-cp 显示 [[双侧额部]bod|bod-pt 或 [额顶部]bod|bod-pt 有 [蛛网膜下腔]bod|bod-pt 增宽]sym|sym-sn
- ③ 不良反应有 [[气管]bod|bod-pt 痉挛]sym|sym-sd 、 [[心功能]ite 不全]sym|sym-sd 、 [恶心]sym|sym-sd 、 [呕吐]sym|sym-sd 。
- (3) 如果前后文中有“表现为”、“表现有”、“有”、“不良反应有”、“症状有”、“等症状”、“反应”等描述症状出现的词，则标注对应实体为“临床表现”。
- ① 不良反应有 [[气管]bod|bod-pt 痉挛]sym|sym-sd 、 [[心功能]ite 不全]sym|sym-sn 、 [恶心]sym|sym-sd 、 [呕吐]sym|sym-sd 。
- ② [新生儿化脓性脑膜炎]dis|dis-ds 临床表现常不典型，尤其是早产儿，一般表现包括 [面色苍白]sym|sym-sd 、 [反应欠佳]sym|sym-sd 、 [少哭少动]sym|sym-sd 、 [拒乳或吮乳减少]sym|sym-sd 、 [呕吐]sym|sym-sd 和 [发热]sym|sym-sd 或 [[体温]ite 不升]sym|sym-sd 等。特殊表现有：[神志改变]sym|sym-sd ：[烦躁]sym|sym-sd 易激惹、[惊跳]sym|sym-sd 、 [突然尖叫]sym|sym-sd 和 [嗜睡]sym|sym-sd 、 [神萎]sym|sym-sd 等。（“一般表现为、特殊表现有”表明后面对应的实体应标注为“临床表现”）
- ③ [肾脏功能损害]dis|dis-ch 可出现 [蛋白尿]sym|sym-sd 、 [少尿]sym|sym-sd 和 [无尿]sym|sym-sd 等症状。（“等症状”表明前面对应的实体应标注为“临床表现”）
- (4) 对于临床表现中的“症状”，一般都是描述病人的不适感受或痛苦表现等主观感受，并且症状会有修饰成分，表示其**严重程度**、**频率**等，为保证标

注完整性，在标注时应该将修饰和症状一起标注。

- ① 任何原因导致 [小肠病变]dis|dis-ch 均可使 [叶酸]dru 和 [维生素 B12]dru 的 [吸收障碍]sym|sym-sd，从而导致两者的缺乏，如 [慢性腹泻]sym|sym-sd 可严重影响 [叶酸]dru、[维生素 B12]dru 的吸收。（“慢性腹泻”：“慢性”是对“腹泻”这个症状的修饰，故此时需标注“慢性腹泻”）
 - ② 遇紧急情况，[气管]bod|bod-pt 异物导致 [严重呼吸困难]sym|sym-sd 和 [窒息]sym|sym-sd 时，或患儿 [[牙关]bod|bod-pt 紧闭不能张开]sym|sym-sd 者，可考虑 [气管切开]pro|pro-tp 和 [环甲膜穿刺]pro|pro-tp 以改善通气。（“严重呼吸困难”：“严重”是对“呼吸困难”这个症状的修饰，故此时需标注“严重呼吸困难”）
 - ③ [胃肠道疾病]dis|dis-ds 目前公认的定义为 [3 次或反复多次的发作性顽固的恶心和呕吐，每次发作持续数小时至数日，2 次发作间期有长达数周至数日的完全无症状间隙期]sym|sym-sd（临床表现的频率，持续时间应该整体标注）
- (5) 对于临床表现中的“体征”，一般都是通过检查设备观察到的病理或生理改变等客观表现，因此多出现在表示检查的词后面，比如“见”，“可见”、“及”、“闻及”、“未闻及”等，这里有两种情况：一种是“医疗程序/医疗设备+（检查词）+体征”，则在标注时仅将**对应的体征**标注为“临床表现”，不用标注此类检查词；另一种是“身体部位/主体+（检查词）+体征”，为保证标注意义完整，则在标注时应该将“身体部位+（检查词）+体征”作为整体标注成“临床表现”。
- ① [阳性体征]sym|sym-sn 主要表现在 [呼吸系统]bod|bod-mt，多数患儿 [[肺部]bod|bod-pt 可闻及啰音]sym|sym-sn，[啰音]sym|sym-sn 可出现在疾病早期或中期，多为 [细湿啰音]sym|sym-sn，[啰音]sym|sym-sn 出现部位多与 [胸部 X 线片]pro|pro-cp 显示的 [肺实变]sym|sym-sn 部位一致，可以双侧或单侧，随着疾病好转而消失。（“肺部可闻及啰音”属于“身

体部位+检查词+体征”，故应整体标注为“临床表现”）

② 行 [头颅 CT]pro|pro-cp 和 [MRI 检查]pro|pro-cp 显示 **[[双侧额部]bod|bod-pt 或 [额顶部]bod|bod-pt 有 [蛛网膜下腔]bod|bod-pt 增宽]sym|sym-sn**（“头颅 CT 和 MRI 检查显示双侧额部或额顶部有蛛网膜下腔增宽”属于“医疗程序+检查词+体征”，故只需将“双侧额部或额顶部有蛛网膜下腔增宽”标注为“临床表现”）

③ **[[心尖]bod|bod-pt 上 翘 明显]sym|sym-sn**，[侧位片]pro|pro-cp **[[心脏前缘]bod|bod-pt 向 [胸骨]bod|bod-pt 方向 扩大]sym|sym-sn**。

（“侧位片心脏前缘向胸骨方向扩大”属于“医疗程序+体征”，故只需将“心脏前缘向胸骨方向扩大”标注为“临床表现”）

(6) 描述“临床表现”时，通常是对病人（一个或多个）身体部位进行描述，在标注时应该与这些身体部位一起标注。若出现“部位/主体+有/无+临床表现”，应该整体标注为“临床表现”。若出现“无+症状/体征”，“无”作为描述临床表现的一种修饰成分，应该整体标注为“临床表现”。

① 凡 病程 在 48 小时 以内 ， **[[腹部]bod|bod-pt 无 肠坏死 征象]sym|sym-sn** 者 均 可 采用 [空气灌肠]pro|pro-tp 诊 断 和 治 疗 。（“腹部无肠坏死征象”：“无”这个否定词将部位和症状分隔开，应当整体标注）

② [尿路梗阻]dis|dis-ds 最 常 见 部 位 为 [后尿道瓣膜]bod|bod-pt 处 ， 其 占 男 性 新 生 儿 [尿路梗阻]dis|dis-ds 的 50% ， 常 伴 有 **[[膀胱逼尿肌]bod|bod-pt 无 抑制性 收缩]sym|sym-sn** ， 其中 25% 患 儿 有 [尿失禁]sym|sym-sd 。（“膀胱逼尿肌无抑制性收缩”：“无”这个否定词将部位和症状分隔开，应当整体标注）

③ [发热]sym|sym-sd 伴 **[[面]bod|bod-pt 、 [颈]bod|bod-pt 及 [上胸部]bod|bod-pt 潮红]sym|sym-sd**。（“面、颈及上胸部潮红”多个部位的症状应该整体标注）

④ 行 [头颅 CT]pro|pro-cp 和 [MRI 检查]pro|pro-cp 显示 **[[双侧额部]bod|bod-pt 或 [额顶部]bod|bod-pt 有 [蛛网膜下腔]bod|bod-pt 增宽]sym|sym-sn**（“双侧额部或额顶部有蛛网膜下腔增宽”多个部位

的症状应该整体标注)

- ⑤ 轻型：**[无脱水]sym|sym-sd**，**[无[中毒]dis|dis-ip 症状]sym|sym-sd**；中型：轻度 1~3 天，常伴**[发热]sym|sym-sd**和**[[上呼吸道感染]dis|dis-ch 症状]sym|sym-sd**。

(7) 表示症状出现的词，不应包括在症状里，比如发作、出现等。

- ① 热退后出现**[低血压]sym|sym-sd**、**[休克]sym|sym-sd**、**[少尿]sym|sym-sd**，继而出现**[血尿]sym|sym-sd**应考虑**[流行性出血热]dis|dis-ds**。（“出现”不应该被标注）
- ② 青春期左右发病，7~17 岁起病，发病年龄高峰在 10~12 岁，男女性别无差异，**[失神]sym|sym-sd**发作频率较少，不一定每天均有发作，多伴有**[[全身]bod|bod-pt 强直 - 阵挛]sym|sym-sd**发作。（“发作”不应该被标注）

(8) 若上下文有指出症状有哪些、是什么，则对应的实体标注为“临床表现”，并且如果临床表现实体后面紧跟“症状”、“征象”、“体征”、“反应”，此时应该将实体和此类词语整体标注；若是非临床表现实体后面紧跟此类词语，则不标注；若单独出现“症状”、“体征”、“征象”、“反应”表示的是一种统称含义，则不标注。

- ① **[真菌性肺炎]dis|dis-ds**的症状和体征无特异性（“症状”和“体征”不标注）
- ② **[X 线胸片]pro|pro-cp**上已经可以出现**[[两肺]bod|bod-pt 细颗粒影]sym|sym-sn**，为**[[间质]bod|bod-mt 水肿]sym|sym-sn**的征象，并可能与临床状况并不一致。（“两肺细颗粒影、间质水肿”标注为“症状”；“征象”不标注）
- ③ **[[中毒]dis|dis-ip 症状]sym|sym-sd**与**[[颅高压]dis|dis-ds 征象]sym|sym-sn**明显、**[[神经系统]bod|bod-mt 局灶定位体征]sym|sym-sn**出现，**[神经影像学检查]pro|pro-cp**帮助诊断。
（“中毒症状、颅高压征象、神经系统局灶定位体征”应该分别作为整体标注）
- ④ **[关节]bod|bod-pt**症状多表现为**[肿胀]sym|sym-sd**、**[疼**

痛]sym|sym-sd 、 [发热]sym|sym-sd 、 [触痛]sym|sym-sd 及 [活动障碍]sym|sym-sd 。（“关节症状”虽然有“症状”后缀，但是“关节”是身体部位，表述的含义并非是一种具体的临床表现，故不标注为“临床表现”）

⑤ 主要 不良 反应 为 [[粒细胞]bod|bod-mt 减少]sym|sym-sn 、 [关节炎]dis|dis-ds 、 [肌肉痛]sym|sym-sd 等 ， 少 部分 病 例 出 现 [[胃肠道]bod|bod-pt 反应]sym|sym-sd

⑥ [无 [中毒]dis|dis-ip 症状]sym|sym-sd ； 中 型 ： 轻 度 1 ~ 3 天 ， 常 伴 [发热]sym|sym-sd 和 [[上呼吸道感染]dis|dis-ch 症状]sym|sym-sd 。（“无中毒症状、上呼吸道感染症状”分别整体标注为“临床表现”，“中毒、上呼吸道感染”应该分别作为“疾病”嵌套标注在内）

(9) 综上所述，可以发现在标注“临床表现”这一实体时，可能是一个词、一个短语，甚至可能是一句话。考虑到标注含义的完整性和可理解性，当多个标注实体是相互独立，即单独的实体也是一种“临床表现”时，标注时应分开标注；当多个以标点符号间隔开的标注实体是不独立的，即描述的是一个身体部位/主体的多种临床表现、或多个身体部位/主体的临床表现，若分开标注，则失去标注含义的完整性，则此时应该作为整体标注。另外，疾病的危害或产生的后果不应该标注“临床表现”。

① [HSP]dis|dis-ds 病理 变化 为 广 泛 的 [白细胞碎裂性小血管炎]dis|dis-ds ， 以 [毛细血管炎]dis|dis-ds 为主 ， 亦 可 波 及 [静脉]bod|bod-pt 和 [小动脉]bod|bod-pt ； [[血管壁]bod|bod-mt 可见 [胶原纤维]bod|bod-mt 肿胀 和 坏 死 ， [中性粒细胞]bod|bod-mt 浸润，周围 有 散 在 核 碎 片]sym|sym-sn ； [[间质]bod|bod-mt 水 肿 ， 浆 液 性 渗 出 ， 同 时 可 见 渗 出 的 [红细胞]bod|bod-mt]sym|sym-sn ； [[内皮细胞]bod|bod-mt 肿胀 ， 可 有 血 栓 形 成]sym|sym-sn 。（“胶原纤维肿胀和坏死，中性粒细胞浸润，周围有散在核碎片”这些都是对血管壁的描述（即一个身体部位的多种临床表现），若分开标注则失去了标注的完整性，故应该和血管壁合并作

为整体标注为“临床表现”；同理，“间质水肿，浆液性渗出，同时可见渗出的红细胞”和“内皮细胞肿胀，可有血栓形成”亦应该整体标注）

② [发热]sym|sym-sd 伴 [[面]bod|bod-pt 、 [颈]bod|bod-pt 及 [上胸部]bod|bod-pt 潮红]sym|sym-sd 。（“面、颈及上胸部潮红”描述的是多个部位的临床表现，应该整体标注）

③ [起病 急]sym|sym-sd ， 常 [发热]sym|sym-sd ， [[大便]bod|bod-mt 稀水 样 ， 量 多 ， 腥臭味 可 不 明显]sym|sym-sd 。（“大便稀水样，量多，腥臭味可不明显”描述的是一个主体的多种临床表现，应该整体标注）

④ 但 可 有 以 下 特 点 ： [持续 高热 不 退]sym|sym-sd ； [明显 烦躁不安]sym|sym-sd ； [咳痰 无色 透明 、 黏稠]sym|sym-sd ；（“咳痰无色透明、粘稠”描述的是一个主体的多种临床表现，应该整体标注）

(10) 临床上的“指征”，一般是指手术指征。所谓手术指征，也就是手术适应症，是指当某种疾病符合诊疗常规所规定的标准，采用非手术治疗方式无法治愈疾病，采用手术方式将有助于疾病的治疗时，所应采用的手术方式。任何一项应予手术的疾病都有一定的标准，比如腰椎管狭窄症手术指征是：活动后腰及腿痛，影响生活工作，经保守治疗不愈者，进行性跛行加重，或站立时间渐缩短者。故在标注工作中，标注者应该将对应的指征标注为“临床表现”。

① [输血]pro|pro-tp 指征： [[心率]jite > 110 次 / 分]sym|sym-sn ；
[[红细胞]jite < $3 \times 10^{12} / L$]sym|sym-sn ； [[血红蛋白]jite < 70 g /
L]sym|sym-sn ； [[收缩压]jite < 12 kPa (90 mmHg)]sym|sym-
sn 。

② 国内 专家 提出 大约 70% 的 [腹泻病]dis|dis-ds 不 需要 也 不
应该 用 [抗生素]dru 。 临床 指征 为 ： ① [血便]sym|sym-sd ；
② 有 [里急后重]sym|sym-sd ； ③ [大便镜检]pro|pro-cp [[白细
胞]bod|bod-mt 满 视野]sym|sym-sn ； ④ [[大便 pH]jite 7 以

上]sym|sym-sn 。

5.3 医疗程序

“医疗程序”泛指为诊断或治疗所采取的措施、方法及过程，包括检查程序和预防或治疗程序。“医疗程序”（tag 为 pro）实体分为检查程序（tag 为 pro|pro-cp）和治疗或预防程序（tag 为 pro|pro-tp）两个子类。

5.3.1 检查程序

检查程序包括检查方法、专项检查、医学影像检查等，检查方法是医生为达到化验目的而采取的某种手段，如尿液分析、活检、穿刺等；专项检查是病人通常情况下做的某种检查，包括体检、孕前检查、妇科专项检查、男子专项检查、盆腔检查、产检等；医学影像检查是放射科或核医学部门的一项程序，主要包括 X 光、超声波、CT（计算机断层扫描）、MRI（核磁共振成像）、乳房照相、荧光检查、PET（正电子发射扫描技术）扫描和骨扫描等。此类医学实体的二级分类标签为 pro|pro-cp。

- ① [两肺听诊]pro|pro-cp 可闻及 [干湿性 啰音]sym|sym-sn ， [胸部 X 线片]pro|pro-cp [[两肺]bod|bod-pt 有 斑片状 模糊 阴影]sym|sym-sn
- ① [肝活检]pro|pro-cp 应争取在起病后 4 ~ 5 日内进行
- ② 婴幼儿目前尚无统一的 [支气管激发试验]pro|pro-cp 或 [舒张试验]pro|pro-cp 的方法和标准
- ③ 如对本症病人在青春期后作 [睾丸活体组织检查]pro|pro-cp
- ④ 常在出生后 3 ~ 4 周（或孕周龄 32 ~ 34 周）第一次做 [眼底检查]pro|pro-cp ，采用 [暗室散瞳]pro|pro-cp 后做 [双眼间接检眼镜检查]pro|pro-cp ，每 2 周复检 1 次
- ⑤ 尽管 [右位心]sym|sym-sn 可通过 [体格检查]pro|pro-cp 发现 ，但通常需经 [胸部 X 线平片]pro|pro-cp 进行诊断 。

5.3.2 治疗或预防程序

治疗或预防程序是医生为达到治疗目的而采取的某种手段，如化疗、放疗、

手术、透析、紧急救治等。此类医学实体的二级分类标签为 **pro|pro-tp**。

- ① **[静脉注射]pro|pro-tp** 用 **[丙种球蛋白]dru** (**[IVIG]dru**) 对部分 **[狼疮]dis|dis-ds** 患儿有一定疗效
- ② **[骨髓移植]pro|pro-tp** 及 **[肾移植]pro|pro-tp** 后均可发生 **[HUS]dis|dis-ds**，发生率分别为 3.4% 和 6% ~ 9%。一旦发生 **[骨髓移植]pro|pro-tp** 后 **[HUS]dis|dis-ds**，预后凶险，可能与大剂量 **[化疗]pro|pro-tp**、**[放疗]pro|pro-tp**、**[排异反应]sym|sym-sd**、**[感染]sym|sym-sd** 等有关。
- ③ 如因特殊情况如 **[急诊检查]pro|pro-cp** 和 **[手术治疗]pro|pro-tp** 等，需要搬动病人，应保持 **[头部]bod|bod-pt** 固定。
- ④ 近年来，采用大部分 **[脾栓塞疗法]pro|pro-tp** 治疗 **[HS]dis|dis-ds** 有一定近期疗效，对 5 岁以前患儿应用该疗法，可以避免 **[免疫功能的严重下降]sym|sym-sd**。但这部分病例将来是否必要作 **[脾切除]pro|pro-tp** 尚无定论。
- ⑤ 对 **[贫血]sym|sd** 较重者，需 **[输血]pro|pro-tp**。

注意：

- (1) “医疗程序”实体内部不存在分词，可包含标点符号。
- (2) 当医疗程序前面有身体部位，即表示对某部位进行检查或治疗，应该“部位+医疗程序”整体标注。
 - ① **[脑 CT]pro|pro-cp** 或 **[磁共振成像]pro|pro-cp** (**[MRI]pro|pro-cp**) 检查可发现结构异常，但这些检查大多限于临床研究所用。(“脑 CT”应该整体标注)
 - ② 尽管 **[右位心]sym|sym-sn** 可通过 **[体格检查]pro|pro-cp** 发现，但通常需经 **[胸部 X 线平片]pro|pro-cp** 进行诊断。(“胸部 X 线平片”应该整体标注)
 - ③ **[大便镜检]pro|pro-cp** **[[白细胞]bod|bod-mt 满视野]sym|sym-sn**；
(“大便镜检”应该整体标注)
- (3) 辅助治疗和非药物治疗也标注为“医疗程序”。
 - ① 主要为 **[控制体重]pro|pro-tp**、**[降低体质指数]pro|pro-tp**；**[注**

意生活规律]pro|pro-tp， [提倡体育锻炼]pro|pro-tp

② [疼痛缓解]pro|pro-tp 与 [姑息治疗]pro|pro-tp 策略

③ 小于 5 mm 没有临床症状的病人可仅以 [药物保守治疗]pro|pro-tp

(4) 当医疗程序作为连续性的动作出现时，应该将整个具体过程整体标注为“医疗程序”，实体内部可以包含标点符号。

① 通过 [调节潮气量、触发/频率，获得稳定的分钟通气量和高同步化]pro|pro-tp。常用的 [呼吸机]equ|equ-te 治疗参数为：[将潮气设置在 6~8ml/kg（实际潮气乘以体重），通气频率视体重和病情设定，一般 30~40 次/分钟，PIP 在 20~25cmH₂O]pro|pro-tp

② 首先要做的是 [纠正水、电解质平衡紊乱]pro|pro-tp 和 [支持疗法]pro|pro-tp

(5) “X 线”单独出现，标注为“医疗程序”；“紫外线”单独出现不标注，若为“紫外线照射”，则标注为“医疗程序”。

① JIA 早期（病程 1 年左右）[X 线]pro|pro-cp 仅显示 [[软组织]bod|bod-mt 肿胀]sym|sym-sn

② 紫外线和直射阳光可迅速降低 [病毒]mic 活力（“紫外线”单独出现不用标注）

5.4 医疗设备

“医疗设备”泛指为诊断或治疗所使用的工具、器具、仪器或机器等，包括检查设备、治疗设备。“医疗设备”（tag 为 equ）实体分为检查设备（tag 为 equ|equ-ce）和治疗设备（tag 为 equ|equ-tp）两个子类。

5.4.1 检查设备

检查设备通常指的是医院中用来检查或检验的仪器，比如血细胞分析仪、生化分析仪、生理记录仪等。此类医学实体的二级分类标签为 equ|equ-ce。

① 通过 [血细胞分离仪]equ|equ-ce 可分离得 [白细胞]bod|bod-mt

② 并将它们的尾部分别连接于 [生理记录仪]equ|equ-ce 和 [程序

刺激仪]equ|equ-ce 。

5.4.2 治疗设备

治疗设备是医生为达到治疗目的而单独或者组合使用于人体的仪器、设备、器具，如注射器、供养面罩、呼吸器等。此类医学实体的二级分类标签为 equ|equ-te。

- ① 由于 医用 器械（如 [吸痰器]equ|equ-te 、 [雾化器]equ|equ-te 、 [供氧面罩]equ|equ-te 、 [气管插管]equ|equ-te 等）消毒不严，暖箱湿度过高使 [水生菌]mic 易于繁殖，或使用 [呼吸机]equ|equ-te 时间过长等引起 [肺炎]dis|dis-ds
- ② [氧气源]equ|equ-te 、 [给氧鼻导管]equ|equ-te 、 [简易呼吸器]equ|equ-te （ [面罩]equ|equ-te 、 [复苏皮囊]equ|equ-te ） 、 [气管插管导管]equ|equ-te 、 [喉镜]equ|equ-te 、 [中心静脉及外周静脉导管]equ|equ-te 、 [留置针]equ|equ-te 等。
- ③ [血管内支架]equ|equ-te 在 [先天性心脏病]dis|dis-ds 中的应用：常用 [4 通道测压导管]equ|equ-te ，可测不同部位及不同方向的压力。
- ④ [输液泵]equ|equ-te 包括注射装置、机械活动装置、电动装置和电子控制装置，适用于新生儿转运的 [注射输液泵]equ|equ-te 应符合下列要求：适用的 [注射器]equ|equ-te 范围大，包括 10 ~ 50 ml [注射器]equ|equ-te 均可适用。

注意：

- (1) “医疗设备”实体内部不存在分词，可包含标点符号。
- (2) 医疗设备的属性不可标注为“医疗设备”。
 - ① 适用的 [注射器]equ|equ-te 范围大，包括 10 ~ 50 ml [注射器]equ|equ-te 均可适用。（其中“10 ~ 50 ml”是注射器的属性不应该标注为“医疗设备”）
- (3) 特殊情况：“氧气”的标注问题。当氧气作为一种治疗、急救的物品时，应该标注为“医疗设备”。

- ① [超声心动图]pro|pro-cp 室内 配备 [氧气]equ|equ-te 、 [吸引器]equ|equ-te 及 [急救药物]dru 可 增加 安全性。

5.5 药物

药物广义上是指用来预防、治疗及诊断疾病的物质，其中有激素类药物、非激素类药物。激素类药物以人体或动物激素（包括与激素结构、作用原理相同的有机物）为有效成分的药物，如雄性激素、胰岛素、生长激素；除此之外就是非激素类药物，这里不做解释，如苯唑西林、红霉素、阿莫西林、头孢、抗生素、维生素等，另外药物也包括临床诊断试剂。

- ① 怀疑 [金葡菌]mic 者 宜 首选 [苯唑西林]dru 或 [万古霉素]dru ； [厌氧菌感染]dis|dis-ds 给予 [青霉素 G]dru 、 [克林霉素]dru 、 [哌拉西林钠]dru 、 [甲硝唑]dru 等 。
- ② 常用 的 药物 有 ； [维生素 E]dru 、 [肌苷]dru 、 [三磷酸腺苷]dru 以及 [中药]dru 等 。 利用 [肾上腺皮质激素]dru 和 [联苯双酯]dru 等 可 降低 [血清酶]bod|bod-mt 水平 。 有人 用 [别嘌呤醇]dru 治疗 本病 ， 患者 的 临床 症状 有所 好转 ， [[血清 CK]ite 下降]sym|sym-sn 。 有人 提出 早期 给予 [乳酸钠]dru ， 可 增强 患者 的 肌力 。
- ③ 多数 [革兰阴性杆菌]mic 耐药率 较高 ， 一旦 诊断 此类 [感染]sym|sym-sd ， 宜 首选 第三代 [头孢霉素]dru 或 [复合 β -内酰胺]dru 类 （ 含 [β -内酰胺酶抑制剂]dru ） 。 如 致病 菌株 产生 [超广谱 β -内酰胺酶]mic （ [ESBL]mic ） ， 应 选用 [头孢霉素]dru 类、 [复合 β -内酰胺]dru 类 ， 严重 者 选用 [碳青霉烯类抗生素]dru 如 [亚胺培南]dru 。
- ④ 目前 应用 [α -干扰素]dru 治疗 [HBV-GN]dis|dis-ds 取得 明显 效果 ， 且 远期 作用 尤 佳 ， 剂量 为 [α -干扰素]dru 20 万 U/（ kg \cdot 次 ） 。
- ⑤ 纠正 [脱水]sym|sym-sd ： 轻 、 [中度脱水]sym|sym-sd ， 在 最初 2 小时 内 [静滴]pro|pro-cp 5% ~ 10% [葡萄糖生理盐水]dru 20 ~ 40 ml /kg 。 ② 纠正 [低血钠]dis|dis-ds ： 补钠量 （ mmol/L ） 按 （ 135 - 测得值 ） \times 0.6 \times 体重 计算 ， 初 8 ~ 12 小时 给予 总量 的 一

半，余半量放入维持量中补给；可用 [9 α -氟氢可的松]dru 0.05 ~ 0.1 mg / d 口服。③ 纠正严重 [高血钾]dis|dis-ds：按 [葡萄糖]dru 0.5 g / kg 加 [胰岛素]dru 0.3 U / kg [静滴]pro|pro-tp。④ 补充 [HC]dru 100 ~ 200 mg / (m²•d) 或 [醋酸可的松]dru 125 ~ 250 mg / (m²•d)，分 3 次口服，1 周后减量，3 ~ 4 周后减至维持量。

⑥ [染色体]bod|bod-mt 在正常情况下呈杆状，经 [秋水仙素]dru 处理后使原来已纵裂的 [染色体]bod|bod-mt 不能分开。

注意:

- (1) “药物”实体内部不存在分词，可包含标点符号。
- (2) 药物的属性不可标注为“药物”。
 - ① 可用 [9 α -氟氢可的松]dru 0.05 ~ 0.1 mg / d 口服（其中“0.05 ~ 0.1 mg / d”是药物的属性不应该标注为“药物”）
- (3) 上下文有关键的提示文字的治疗：“治疗”、“化疗”、“放疗”、“口服”、“注射”、“可补充”、“药物”等，其对应的实体标注为治疗。如果标注实体后面紧跟着“治疗”、“化疗”等，应和这些词语作为整体一起作为标注。
 - ① 对于诊断困难或诊断后 [铁剂治疗]pro|pro-tp 效果不理想的患儿，有条件的单位可以考虑进行，以明确或排除诊断。
 - ② [肾上腺皮质激素]dru 为首选药物，且 [多药联合治疗]pro|pro-tp 并不比采用单一药物效果好，但对多 [[脏器]bod|bod-pt 受累]sym|sym-sn 病人应采用 [联合化疗]pro|pro-tp（“肾上腺皮质激素为首选药物”明确指出“肾上腺皮质激素”是一种药物，故应该标注为“药物”；“联合化疗”应该作为整体标注）
 - ③ 纠正严重 [高血钾]dis|dis-ds：按 [葡萄糖]dru 0.5 g / kg 加 [胰岛素]dru 0.3 U / kg [静滴]pro|pro-tp（“静滴葡萄糖加葡萄糖”明确指出“葡萄糖”是一种药物，故应该标注为“药物”）
- (4) 大部分药物的统称均标注，比如营养素、抗菌药物、急救药物等，此类虽然是统称，但是有对应的治疗范畴，所以应该标注。特殊情况：像常用药、

药物等单独出现时，此类统称范围太广，不应标注。

- ① 对营养支持的需要明显不同于成人，对 [营养素]dru 成分的要求更高。
- ② 配备 [急救药物]dru 可增加安全性。

5.6 医学检验项目

检查涉及到的体液检查项目、重要生理指标以及其他检查项目，本文规定“医学检验项目”主要针对人体而言，是能够通过设备或实验检测出的项目，并且是能够被量化，有其对应的测量值或指标值。

- ① 做好监护对评价患儿病情、指导治疗和判断预后具有积极意义。基本的监护包括神志、[心率]ite、[脉搏]ite、呼吸、[血压]ite、[体温]ite、[尿量]ite、[血乳酸含量测定]pro|pro-cp 和 [血气分析]pro|pro-cp
- ② 等，也可根据情况选用测 [血压]ite、[中心静脉压]ite、[肺动脉楔压]ite、[胃肠黏膜内 pH 值]ite 和 [超声心动图监测]pro|pro-cp 等
- ③ 常用于 [SASP]dru 能耐受、对 [肾上腺皮质激素]dru 依赖及病变广泛不能手术者。应用时需定期检查 [白细胞计数]ite 及 [血小板计数]ite。
- ④ 伴有轻度或重度 [高血压]dis|dis-ds 的患儿可能为数不少，但由于往往忽略测量婴幼儿 [血压]ite，故报道者不多。

注意：

- (1) “医学检验项目”实体内部不存在分词，可包含标点符号。
- (2) 测量值不应该标注为“医学检验项目”。
 - ① [输血]pro|pro-tp 指征：[[心率]ite > 110 次 / 分]sym|sym-sn；
[[红细胞]ite < $3 \times 10^{12} / L$]sym|sym-sn；[[血红蛋白]ite < 70 g / L]sym|sym-sn；[[收缩压]ite < 12 kPa (90 mmHg)]sym|sym-sn。
(“>110 次/分、<3×10¹²/L”等均不标注)
- (3) 上下文有关键的提示文字的检查：“检查”、“检”、“查”、“提示”、

“示”、“查体”、“试验”、“测定”、“听诊”、“观察”、“诊断”等，其对应的检查项目标注为“医学检验项目”。

① 对 [腹泻]sym|sym-sd 较重的患儿，应及时检查 [血 pH]ite、[二氧化碳结合力]ite、[碳酸氢根]ite、[血钠]ite、[血钾]ite、[血氯]ite 及 [血渗透压]ite，对于诊断及治疗均有重要意义。（前面有“应及时检查”，后面对应的实体应标注为“医学检验项目”）

② 密切观察病人的意识、瞳孔、[血压]ite、[呼吸]ite、[脉搏]ite 及 [体温]ite 的变化，掌握病情发展的动态，及时采取措施。病情稳定者要及时进行原因检查和治疗。

（“密切观察…的变化”：后面对应的实体应标注为“医学检验项目”）

(4) 在病人的临床表现中，有时会涉及到对医学检验项目的描述，比如一些生理指标等，标注时应该将这些生理指标嵌套标注为“医学检验项目”。

① 表现为 [血尿]sym|sym-sd、[蛋白尿]sym|sym-sd、[水肿]sym|sym-sd、[[尿量]ite 减少]sym|sym-sd 等。（“尿量减少”：“尿量”在此处表明是一项生理指标，故应标注为“医学检验项目”）

② 对分娩时 [急性胎母大量失血]sym|sym-sd 所致的 [低 [血容量]ite 休克]sym|sym-sd，需立即给 [生理盐水]dru 扩充 [血容量]ite，（“低血容量休克”：“血容量”在此处表明是一项生理指标，故应标注为“医学检验项目”）

③ 患者的临床症状有所好转，[[血清 CK]ite 下降]sym|sym-sn

④ [[肺动脉毛细血管压]ite > 12 mmHg]sym|sym-sn 时，[[上下叶]bod|bod-pt 血管影相当或 [上叶]bod|bod-pt 略有增多]sym|sym-sn

⑤ [[肺毛细血管压]ite 超过 25 mmHg]sym|sym-sn 时，[[肺泡]bod|bod-pt 水肿形成蝶翼征]sym|sym-sn。

⑥ 如反复 [输血]pro|pro-tp > 100 次，[[血清铁饱和度]ite > 80%]sym|sym-sn 者应给予 [螯合剂去铁胺]dru

([deferoxamine]dru) 。

(5) 常规的检查项目：尿常规、血常规、血培养、尿细菌培养等标本采集可获取一系列数据，此类实体应该标注为“医疗检验项目”。

① 结合 [尿常规]ite 及 [尿培养菌落计数]ite 可以做出诊断。

② 故有明显 [感染 [中毒]dis|dis-ip 症状]sym|sym-sd 的患者，应做 [血培养]ite 。

5.7 身体

“身体”身体泛指细胞、组织、及位于人体特定区域的由细小物质成分组合而成的结构、器官、系统、肢体，另外包括身体产生或解剖身体产生的物质等，包括身体部位和身体物质。“身体”(tag 为 bod)实体分为身体部位(tag 为 bod|bod-pt)和身体物质 (tag 为 bod|bod-mt) 两个子类。

5.7.1 身体部位

身体部位包括器官或器官组成、身体系统、身体位置或区域。器官或器官组成是指人体器官或器官组成，比如头、眼球、胳膊、眼睑等；身体系统包括运动系统、消化系统、呼吸系统、泌尿系统、生殖系统等；身体位置或区域是指身体的某个位置，如腹部、下颌、腋窝等。此类医学实体的二级分类标签为 bod|bod-pt。

- ① 生命体征及 [心]bod|bod-pt 、 [脑]bod|bod-pt 、 [肾]bod|bod-pt 等主要器官功能受损
- ② [肾脏]bod|bod-pt 是毒物和毒物代谢产物·泄的主要器官
- ③ 常见于尿布包裹处及光滑 [皮肤]bod|bod-pt 相互直接摩擦部位，如 [肛周]bod|bod-pt 、 [臀部]bod|bod-pt 、 [外阴]bod|bod-pt 、 [腹股沟]bod|bod-pt 以及 [腋窝]bod|bod-pt 、 [颈前]bod|bod-pt 、 [下颌]bod|bod-pt 等处 [皮肤]bod|bod-pt 。
- ④ [消化系统]bod|bod-pt 症状：发病初期主要表现为 [消化系统]bod|bod-pt 症状

5.7.2 身体物质

身体物质可由组织、细胞（包括细胞成分、细胞结构）、生物大分子、以及身体或解剖身体产生的物质。组织包括血液、蛛网膜、眼角膜等；细胞包括红细胞、白细胞、染色体等；生物大分子指生物体细胞内存在的糖类、核酸、蛋白质等，如 DNA、酶、抗体、激素等；身体或解剖身体产生的物质包括唾液、精液、粪便等。此类医学实体的二级分类标签为 bod|bod-mt。

- ① [转移性骨肿瘤]dis|dis-ds 的产生主要通过 [淋巴]bod|bod-mt 或 [血液]bod|bod-mt 两条途径，[原发性肿瘤]dis|dis-ds 可以来自全身各处。
- ② [重型肝炎]dis|dis-ds 是 [[肝细胞]bod|bod-mt 发生大量坏死]sym|sym-sn 而陷入 [肝衰竭]dis|dis-ch 的过程，[肝衰竭]dis|dis-ch 能否逆转取决于 [肝细胞]bod|bod-mt 存活的数量。
- ③ 主要依靠流行病学特点及临床特征，遇有可疑病例必须仔细检查 [粪便]bod|bod-mt、[痰液]bod|bod-mt 的 [虫卵]mic。在流行地区对临床症状不明显的患者，应做 [皮内试验]pro|pro-cp 和 [补体结合试验]pro|pro-cp 以助诊断。
- ④ [甲状腺激素]bod|bod-mt 可影响 [脑神经细胞]bod|bod-mt 的生长、迁移和 [树突]bod|bod-mt 的发育。
- ⑤ [生长激素]bod|bod-mt 由 [腺垂体嗜酸性粒细胞]bod|bod-mt 分泌
- ⑥ 有报告用 [X 线]pro|pro-cp 照射全身 [淋巴结]bod|bod-mt（20 Gy / 4 ~ 6 周），可使部分病人取得一定疗效，[肌酐]bod|bod-mt 清除率好转，[dsDNA 抗体]bod|bod-mt 减少，甚或停用 [泼尼松]dru。
- ⑦ [病原学检查]pro|pro-cp 要对 [痰]bod|bod-mt、[胸腔积液]bod|bod-mt、[经皮肺穿刺物]bod|bod-mt 和 [肺泡]bod|bod-mt 灌洗液进行涂片

注意：

- (1) “身体”实体内部不存在分词，可包含标点符号。
- (2) 表示身体部位具体位置的方位词，如“上”、“下”、“左”、“右”、

“部”、“侧”等应当一同标注。

① Ortega 对 IRS I 和 II 的 56 个 [脊柱旁横纹肌肉瘤]dis|dis-ds 的患者作过总结，大部分发生在 [腰部]bod|bod-pt 和 [胸腰部]bod|bod-pt，[肿瘤]dis|dis-ds 直径大于 5 cm，为未分化型和腺泡型。（“腰部、胸腰部”应整体标注）

② [肾盂]bod|bod-pt 朝向 [背侧]bod|bod-pt，与正常 [肾]bod|bod-pt 之下极相融合

(3) 当出现多个身体部位的组合，考虑到分开标注会失去原本含义的情况下，应作为整体标注为“身体”。

① [指、趾甲]bod|bod-pt（“指、趾甲”表述的含义应该是“指甲和趾甲”，若分开标注则失去原本含义，故应该整体标注）

(4) 当出现微生物和身体物质的组合物质时，均标注为身体。

① 或在 该处 形成 [局部 血栓]sym|sym-sn，并释放出 [细菌栓子]bod|bod-mt 到 [血液]bod|bod-mt 循环中（“细菌栓子”作为微生物和身体物质的组合物质，应标注为“身体”）

(5) “补体、抗体”若以身体物质中的组成成分出现时，应该标注为“身体”

① 因为 [脑脊液]bod|bod-mt 中的 [补体]bod|bod-mt、[抗体]bod|bod-mt 浓度明显低于 [血液]bod|bod-mt 循环

(6) 食物的“糖类”、“蛋白质”等不标注为“身体”，标注应该联系上下文判断是否为身体内物质。

(7) 病变细胞依旧存在与身体内部，故仍标注为“身体”，如[肿瘤细胞]bod|bod-mt、[白血病细胞]bod|bod-mt、[狼疮细胞]bod|bod-mt

(8) “基因”、“染色体”应该标注为“身体”。

① 本病是 [常染色体隐性遗传病]dis|dis-ds，[α -L-艾杜糖醛酸苷酶基因]bod|bod-mt 已被鉴定，位于 [染色体 4p16.3]bod|bod-mt，有 14 个 [外显子]bod|bod-mt，并在其中发现了不少 [[基因]bod|bod-mt 突变]sym|sym-sn。

② 嵌合型，核型为 45，[X]bod|bod-mt/46，[XX]bod|bod-mt，约占该病的 25%。[XX]bod|bod-mt 为主的个体临床症状

较轻。（“X、XX”是“X染色体、XX染色体”的代称，故应该标注为“身体”）

5.8 科室

“科室”主要是指医院或医疗机构所设有的部门以及科室，作为检查和治疗中涉及的一个重要的角色。

- ① 对儿童 [肿瘤]dis|dis-ds 必须有多专业的协调工作才能按计划实施诊治计划，因此应建立一个常规的多专业（[外科]dep、[病理科]dep、[血液/肿瘤科]dep、[影像学科]dep和[放疗科]dep为基本组成单位）讨论制度。
- ② [重症监护病房]dep是指用现代医疗设备装备起来，能对各种危重患者的生命活动进行连续、系统的观察。
- ③ 在各级医院、[期末期病房]dep、社区医疗中心

注意：

“科室”实体内部不存在分词，可包含标点符号。

5.9 微生物类

微生物包括细菌、病毒、真菌以及一些小型的原生生物、显微藻类等在内的一大类生物群体，它个体微小，与人类关系密切。微生物亦可分为致病微生物和非致病性微生物。致病微生物是能够引起人生病的致病因素之一，比如流感病毒可以起人类、禽类的流感。此外，致病菌株产生的毒素或酶也划分为微生物类，包括溶血毒素、杀白细胞毒素、肠毒素等。但还存在非致病性甚至有益性微生物，广泛存在自然界、人体皮肤粘膜表面等，例如人与动物肠道内的双歧杆菌、乳酸杆菌，豆科植物的根瘤菌，海水养殖环境的枯草杆菌、光合细菌、硝化细菌等。

致病微生物和非致病微生物均标注为“微生物类”实体。

- ① [鼠伤寒沙门菌]mic也是[流行性腹泻]dis|dis-ds的重要病原，[鼠伤寒菌]mic分布广泛
- ② 其他[细菌]mic也可产生[肠毒素]mic，如[耶尔森菌]mic、[鼠伤

寒沙门菌]mic 、 [金黄色葡萄球菌]mic 、 [空肠弯曲菌]mic 等 。

- ③ 国际上所指的 [CROTCHS 或 TORCH 感染]dis|dis-ds （ 即 [巨细胞病毒]mic 、 [风疹病毒]mic 、 [弓形虫]mic 、 [柯萨奇]mic 和其他 [肠道病毒]mic、 [单纯疱疹]dis|dis-ds 和 [乙肝病毒]mic 、 [HIV]dis|dis-ds 以及其他 [病毒]mic ） 均可 为 新生儿 [肝炎]dis|dis-ds 的 病因 。
- ④ 如 致病 菌株 产生 [超广谱 β -内酰胺酶]mic （ [ESBL]mic ） ， 应 选 用 [头孢霉素]dru 类

注意：

- (1) 标注实体内部不存在分词，可包含标点符号。
- (2) “寄生虫、原虫”应该标注为“微生物类”。
- (3) “病毒、细菌”是“微生物”的子类，故单独出现时均应该标注为“微生物类”。
 - ① 对人和某些动物都可引起疾病，病愈后带菌率又高，因此 [细菌]mic 来源多，发病率高。
 - ② 此法能发现不完整 [病毒]mic
- (4) 出现“病毒+病毒的一部分（基因、DNA）”应该整体标注为“微生物类”。
 - ① 应用 [核酸杂交法]pro|pro-cp 或 [聚合酶链反应]pro|pro-cp （ [PCR]pro|pro-cp ） ， 通过 检测 [病毒特异性核酸]mic 片段 来 发现 相关 [病毒]mic
 - ② 此法能发现不完整 [病毒]mic 如 潜伏 [病毒 DNA]mic

六、实体分类混淆处理

6.1 疾病(dis)和临床表现(sym)

疾病和临床表现的最大区别是：疾病是通过鉴别诊断的，疾病实质上就是身体受损，人体是细胞构成的，身体受损其实就是细胞、细胞组织、细胞群等受损。而临床表现实质上是身体受损后所表现出来的现象，比如说病人的不适感觉、身体出现的异常变化，但是这些往往是病人或者医生看到的表面现象。而作为医生

则需要通过进一步的鉴别诊断来确认病人所患疾病，这也就说明疾病和临床表现存在着本质性的差异。比如“急性肾功能衰竭”这种疾病是多种原因导致肾脏生理功能急剧减低甚至丧失，临床表现有水及电解质失调、代谢性酸中毒等，而医生对于该疾病的诊断，需要通过一系列的检查方可得出结论。

但是，在不同的语境下，相同的词可能会标注不同的实体，比如：高血压、贫血等。“出现”、“导致”等，即可以表明症状，也可以表明疾病。如果前面有具体说明原因，结果导致某种疾病的产生，从而出现什么临床表现，则不同情况下对应的标注实体就是“疾病”或“临床表现”。因此，“临床表现”应该理解为病人主动性的一个动作或者观察到的一种变化，大多数为动词，而“疾病”应该理解为病人被动性的一种获得或者一种结果，大多数为名词。这两者有时会存在歧义，故应该在不同语义下着重分析。

① 临床可分为典型与非典型型。因 **[[11β-OH]bod|bod-mt 缺乏|sym|sym-sn** 而导致 **[[DOC]bod|bod-mt 增加|sym|sym-sn**，可使部分患儿出现 **[高血钠]dis|dis-ds**、**[低血钾]dis|ds-ds**、**[碱中毒]dis|dis-ds** 及 **[高血容量]dis|dis-ds**，导致 **[[高血压]dis|dis-ds 症状|sym|sym-sd**。（“11β-OH 缺乏、DOC 增加”均为病人存在的某些症状，故而导致病人得病；“高血压、碱中毒、高血容量”是病人所得疾病）

问题：“感染”相关的实体标注为“疾病”还是“临床表现”？

- (1) 若出现明确致病原因（病毒、细菌或身体部位等名称）与“感染”组合成词，则整体标注为“疾病”。如 **[HAV 感染]dis**、**[球菌感染]dis**、**[上呼吸道感染]dis**。
- (2) 若不具体指明感染类型的情况下，则一般标注为“临床表现”
- (3) 当单独出现“感染”一词时，若上下文明显表示是对某种疾病的指代，则标注为疾病，否则标注为“临床表现”

综上所述，“感染”一词加修饰词与未加修饰词的区分：若修饰词是明确**致病原因（病毒、细菌或身体部位等名称）**或整体是对某种疾病的指代，则标注为“疾病”；若修饰词是表示**程度或频率**的词且整体不是对某种疾病的指代，则标注为“临床表现”。

6.2 医疗程序(pro)和医疗设备(equ)

医疗程序是指检查过程以及预防或治疗过程，描述的是医生为诊断或者治疗而采取的一系列操作或过程。而医疗设备是指检查设备以及治疗设备，描述的是具体的设备实体（工具、器具、仪器或机器），是医生进行诊断或者治疗的工具。标注者在标注时应该谨慎区分。

- a) 由于 医用 器械（如 **[吸痰器]equ|equ-te**、**[雾化器]equ|equ-te**、**[供氧面罩]equ|equ-te**、**[气管插管]equ|equ-te** 等）消毒不严，暖箱湿度过高使 **[水生菌]mic** 易于繁殖，或使用 **[呼吸机]equ|equ-te** 时间过长等引起 **[肺炎]dis|dis-ds**（“吸痰器、雾化器、供氧面罩、气管插管、暖箱”均应标注为“医疗设备”）
- ② **[血管内支架]equ|equ-te** 在 **[先天性心脏病]dis|dis-ds** 中的应用：常用 **[4 通道测压导管]equ|equ-te**，可测不同部位及不同方向的压力。（“血管内支架、4 通道测压导管”均应标注为“医疗设备”）
- ③ 临床发现有些 **[头颅]bod|bod-mt** 较大的婴儿，行 **[头颅 CT]pro|pro-cp** 和 **[MRI 检查]pro|pro-pc**（“头颅 CT、MRI 检查”均应标注为“医疗程序”）
- ④ 如对本症患者 在 青春期后作 **[睾丸活体组织检查]pro|pro-cp**（“睾丸活体组织检查”均应标注为“医疗程序”）
- ⑤ **[骨髓移植]pro|pro-tp** 及 **[肾移植]pro|pro-tp** 后均可发生 **[HUS]dis|dis-ds**，发生率分别为 3.4% 和 6% ~ 9%。（“骨髓移植、肾移植”应标注为“医疗程序”）

6.3 药物(dru)和身体(bod)

标注实体是某种身体物质时，如生长激素，但是有“口服”、“注射”等字眼显式地表明该实体是一种药物（是外来的），此时应该将该类实体标注为“药物”。否则，如果只是表明该实体在人体中的一种状态（是内在的），应该标注为“身体”。

- ① **[[LH]bod|bod-mt 脉冲式 释放]sym|sym-sn 增加**、**[LH]bod|bod-mt 对 [GnRH]dru 注入 后的 反应 增强 以及 [血]bod|bod-mt|bod-mt 中 [性激素]bod|bod-mt 水平 上升**。（“LH 脉冲式释放增加、LH 对 GnRH

注入后的反应增强”：说明 LH 是体内激素，应该标注为“身体”，GnRH 应该标注为“药物”；“血中性激素水平上升”只是对身体物质所处状态的一种描述，故“性激素”应该标注为“身体”）

- ② [糖尿病]dis|dis-ds 患儿 由于 [[胰岛素]bod|bod-mt 分泌 不足 或 缺如]sym|sym-sn ， 使 [葡萄糖]bod|bod-mt 的 利用 （ 进入 [细胞]bod|bod-mt ） 量 减少 ， 而 增高的 [胰高血糖素]bod|bod-mt 、 [生长激素]bod|bod-mt 和 [皮质醇]bod|bod-mt 等 却 又 促进 [肝糖原]bod|bod-mt 分解 和 [葡萄糖]bod|bod-mt 异生 ， [脂肪]bod|bod-mt 和 [蛋白质]bod|bod-mt 分解 加速 ， 造成 [[血糖]ite 增高]sym|sym-sn 和 [[细胞外液渗透压]ite 增高]sym|sym-sn 、 [[细胞内液]bod|bod-mt 向 [细胞]bod|bod-mt 外 转移]sym|sym-sn 。（“胰岛素分泌不足，葡萄糖利用量减少”只是对身体物质所处状态的一种描述，故“葡萄糖”和“胰岛素”应该标注为“身体”）
- ③ 快速 [静脉]pro|pro-tp 应用 [碳酸氢钠]dru 1 ~ 3 mmol/kg ， 或 [葡萄糖]dru 加 [胰岛素]dru （ 0.5 ~ 1 g [葡萄糖]dru/kg （“静脉葡萄糖加胰岛素”表明是“葡萄糖”和“胰岛素”在这种情况下是药物，故“葡萄糖”和“胰岛素”应该标注为“药物”）

6.4 医学检验项目(ite)和医疗程序(pro)

医学检验项目是指**体液检查项目**、**生理测量**、**重要生理指标**以及其他检查项目，与医疗程序存在一定的区别。如上文所述，当上下文有关键的提示文字的检查：“检查”、“检”、“查”、“提示”、“示”、“查体”、“试验”、“测定”、“听诊”、“观察”、“诊断”、“监护”等，其对应的检查项目标注为“医学检验项目”。但如果标注实体后面紧跟着“检查”、“测定”、“诊断”等，表明是医生为诊断或者治疗而采取的一系列操作或过程，故应当将“医学检验项目”应和这些词语作为整体一起作为标注为“医疗程序”。

- ① 尤其 有 [腹膜炎]dis|dis-ds 或 [肠梗阻]dis|dis-ds 表现者 ， 常规 行 [腹腔穿刺]pro|pro-cp 及 [腹水淀粉酶测定]pro|pro-cp ， 能 提高 早期 诊断率 。（虽然“腹水淀粉酶”是一项医学检验项目，但是后面紧跟“测定”，

则表明是一项医疗程序，故应标注为“医疗程序”)

- ② 做好监护对评价患儿病情、指导治疗和判断预后具有积极意义。基本的监护包括神志、[心率]ite、[脉搏]ite、呼吸、[血压]ite、[体温]ite、[尿量]ite、[血乳酸含量测定]pro|pro-cp 和 [血气分析]pro|pro-cp 等，（虽然“血乳酸含量、血气”是医学检验项目，但是后面紧跟“测定、分析”，则应标注为“医疗程序”）

6.5 医学检验项目(ite)和身体(bod)

标注实体是身体某种物质时，但是在某些检查中涉及到对该类实体存在指标规定，显式地表明该类实体应该是某种检查项目，并且后面通常紧跟着测量值或者指标值，此时应该将该类实体标注为“医学检查项目”。否则，如果只是表明该实体在人体中的一种状态，应该标注为“身体”。

- ① 足月儿生后 6 ~ 12 周时 [血红蛋白 下降至 95 ~ 110 g / L]sym|sym-sn，主要原因有：在 [宫内]bod|bod-pt，胎儿 [血氧饱和度]ite 约 50%，相对缺氧状态使促 [红细胞生成素]bod|bod-mt 含量较高，[红细胞]bod|bod-mt 较多，出生后 [血氧饱和度]ite 显著增高。（“血氧饱和度约 50%、出生后血氧饱和度显著增高”：表明“血氧饱和度”是一种指标，故应该标注为“医学检验项目”；“红细胞生成素含量较高，红细胞较多”只是对身体内物质所处状态的描述，故应该标注为“身体”）
- ② [输血]pro|pro-tp 指征：[[心率]ite > 110 次 / 分]sym|sym-sn；[[红细胞]ite < $3 \times 10^{12} / L$]sym|sym-sn；[[血红蛋白]ite < 70 g / L]sym|sym-sn；[[收缩压]ite < 12 kPa (90 mmHg)]sym|sym-sn。
（“红细胞、血红蛋白”应该是身体物质，但这种情况下，“输血指征... 红细胞<3×10¹²/L、血红蛋白<70g/L”表明“红细胞、血红蛋白”是指标，故应该标注为“医学检验项目”）

七、结束语

本文主要提出了面向统一化的面向医学文本处理的医学实体标注规范，适用于普遍的医学文本，例如医学教材、医学百科、临床路径、病历、医学期刊、检验报告等，并且本文将医学实体重新划分和定义类别，包含“疾病”、“临床表现”、“医疗程序”、“医疗设备”、“药物”、“医学检验项目”、“身体”、“科室”和“微生物类”九大类，分别包含其子类。标注语料的制定，离不开统一的规范。因此，规范的简单性、易操作性和一致性原则对于标注质量起着至关重要的作用。本文正是遵循这些原则，旨在建立一个简易明了的标注规范，文中详细介绍了各个医学命名实体的涵盖范畴，阐述实体间的混淆处理，并且用大量示例举例说明。

后续的工作将从标注数据和数据质量评测两个方面展开。实体标注是未来医学文本处理任务的重中之重，必须保证标注结果的真实可靠，标注者需详读本文制定的标注规范，明确实体含义。标注质量评测方面，采取迭代标注、抽样检查等多种措施，提高标注效率和标注质量。

参考文献

- [1] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* 2004;32(Database issue):D267–D270. doi: 10.1093 /nar/ gkh061
- [2] i2b2 / VA Challenge. Concept Annotation Guidelines. 2010. <https://www.i2b2.org/NLP/Relations/assets/ConceptAnnotationGuideline.pdf>. Accessed 11 Nov.2016.
- [3] Roberts A, Gaizauskas R, Hepple M, Davis N, Demetriou G, Guo Y, et al. The CLEF corpus: semantic annotation of clinical text. *Amia Annu Symp Proc.* 2007;2007:625–9.
- [4] South BR, Shen S, Jones M, Garvin J, Samore MH, Chapman WW, et al. Developing a manually annotated clinical document corpus to identify phenotypic information for inflammatory bowel disease. *BMC Bioinformatics.* 2009;10(S9):S12.
- [5] Lei J, Tang B, Lu X, Gao K, Jiang M, Xu H. A comprehensive study of named entity recognition in Chinese clinical text. *J Am Med Inf Assoc.* 2014;21(5): 808–14.
- [6] Xu, Yan & Wang, Yining & Liu, Tianren & Liu, Jiahua & Fan, Yubo & Qian, Yi & Tsujii, Jun'ichi & Chang, Eric. (2013). Joint segmentation and named entity recognition using dual decomposition in Chinese discharge summaries. *Journal of the American Medical Informatics Association : JAMIA.* 21. 10.1136/amiajnl-2013-001806.
- [7] Yang, Jinfeng & Guan, Yi & He, Bin & Qu, C.-Y & Yu, Q.-B & Liu, Y.-X & Zhao, Y.-J. (2016). Corpus construction for named entities and entity relations on Chinese electronic medical records. *J Softw.* 27. 2725-2746. 10.13328/j.cnki.jos.004880.
- [8] Gao, Yao & Gu, Lei & Wang, Yefeng & Wang, Yandong & Yang, Feng. (2019). Constructing a Chinese electronic medical record corpus for named entity recognition on resident admit notes. *BMC Medical Informatics and Decision Making.* 19. 10.1186/s12911-019-0759-2.
- [9] 俞士汶,段慧明,朱学锋,孙斌. 北京大学现代汉语语料库基本加工规范[J]. 中文信息学报, 2002, 16(5): 51-66.
- [10] 中文引用格式: 杨锦锋,关毅,何彬,曲春燕,于秋滨,刘雅欣,赵永杰.中文电子病历命名实体和实体关系语料库构建.软件学报, 2016,27(11):2725–2746. <http://www.jos.org.cn/1000-9825/4880.htm>