

数据质量问题在大数据环境中的充分暴露





脱离原生系统

系统隐含显示规则 以屏幕显示为目的 以打印输出为目的



多源汇聚

数据口径与业务口径 系统偏差与随机偏差



数据集成

由接口产生的质量问题 词表转化 数据连续性

数据加工

过早标化处理 知识库版本



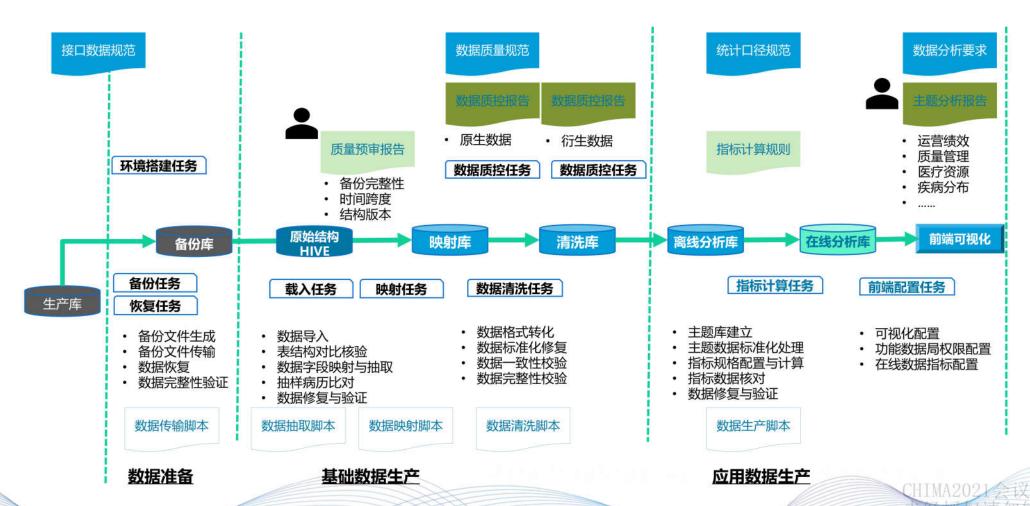


从电子病历分级评价看数据质量评测方法

评价维度	电子病历分级 评价	GB36344	ISO/IEC 25024:2015	电子病历数据质量评估与治理综述研究		
				CIHI 数据质量框架	CHAN K S.Fowles JB	Weiskopf N G Wang C
规范性	✓	✓	✓			
完整性	✓	✓	✓		√	✓
准确性		✓	✓	✓	✓	✓
一致性	✓	✓	✓	✓		✓
时效性	✓	✓		✓		✓
可访问性		✓	✓		✓	
合理性						✓
可比性				✓	✓	
可用性				✓		
可信性			✓			
现实性			✓			
可跟踪性			✓			CHIMA2021会i
可恢复性			✓			未经授权请勿

数据质量改进的PDCA循环





大数据环境中的数据质量评价挑战





	0	I DOOLOOL			
慧药处方	药物名称	与药物通用名称不同时为交(门+位)			
出人無害	总结-出入量mi	單日静脉入量不得超过5000mi(主要指糖盐液体,不包括血液制品)			
出院兒童	推交时间	出荒记象,女核模交时询和记录时揭不可现时为空			
出款小结	W文时间	出禁小组,文档接交的间和记录时间不可同时为空			
非药品质碱	押止时间	出版集者的长期市药基聚磺胺基状态为弊止且有序止时间			
非药品医嘱	医嘱类型	有效状态下,门诊应有检查/检验/处置四类医嘱,住院应有检查/检验/处置/护理/手术/耗材			
患者基本信息		患者扩展号表表住院病案号、住院号、门诊号、医院自制就诊卡号不允许拼接			
患者基本信息	病案号	患者基本信息 当patient_sn在就诊信息表里有就诊类型为住院的Visit时,病案号 不能力空 年龄必须在0-150之内,出生日期幾于1868年			
患者基本信息	年龄				
会诊记录	提交时间	会诊记录,文档提交时间和记录时间不可同时为空			
检查记录	检查状态	影像检查业务表,不存在检查状态为作废和空的数据			
检查记录	检查所见	非作废状态,检查所见和检查结论不可同时为空			
检查记录		检查登记表的主键唯一			
性管に限	学生出版	粉型明测表中参考值范围有值(包含 ^-*)。 高值或值值先值的数据量			
机机记录	伊考尼报	检验明题表中参考镇范围有值(位含 ">" , 信含 ">=") 征信为空; (包含 "=") , 高值为记的数据量			
检验记录	lob_result_id	松后登记者的主城唯一			
检验证明	检验结果	不存在: 混合值、定性、定量都为空的情况			
拉拉定集	定性効果	机耐量蛋白剂 化多甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基甲基			
相称记录	柏柏状态	检验结果业务者。关注检验状态为已作废和空的数据			
检验记录	定性结果	再被定性经验核果值为空			
检验记录	典任信	检验正常信范围有高低,但是高低信全部为空的			
献沙信息	性斯名称	类女各不被位于40%			
保証機能	就设次	统诊类型为住款,计算将一直省总住款次数分布情况:=1,1-10、>11<=20、>20			
献印信息	初度並标志	分布情况			
就设信息	出版时间	出限时间为党的教堂			
製砂塩皂	位院天徽	最大值。最小值。平均值。中位值			
献设位息		就诊信息 往院患者要求有病案号、往院号			
MUSE		联设位组 门设盘者要求有门边号			
就设信息	人政利室	人拉科安加·比拉科安不一样的数量。如果为OSI可能有问题(未来和转科)			
就设信息	人科府河	人能料室与比較料室不一样的數量、如果力O則可能看问题 (未采取转耗) 人與时间举于人能时间的数量。存在則可能看问题 (逻辑错误。应该是先人能离 人料)			
被沙信息		虚拟VID重复			
门急诊发药	数量	门诊发药出现1000次以上,单位为"片 粒 胶囊" 数量小于5的			
门急诊发药		门诊发药表与医嘱表关联			
评分记录	评分项目值,评分项目值单位	评分项目值和评分项目值单位两者均有值			
药品类医嘱	医嘱id	住院草药医嘱Id没有聚合			
药品类医嘱	医嘱状态	住院药品医嘱状态数量,需要关注作度医嘱、没有状态的医嘱			
药品类医嘱		如果存在医嘱开始时间、医嘱结束时间、频率相同而组标识不一致则有问题			
药品类医嘱	医嘱状态				
药品类医嘱		住院非药品医曬状态数量,需要关注作废医嘱、没有状态的医嘱 3.3版本之前住院草药医嘱の(groupid相同,但是因为草药付数、用法、制备方法、频次、总剂量单位、包装数量单位、剂量单位 剂型的原因没有聚合 出版重素酶长期药品医嘱需要状态为停止目有停止时间			
药品类医嘱	停止时间				
药品类医嘱	医嘱名称	住院常用中草药药品,未归入中草药			
药品类医嘱	医嘱名称	住院药品医嘱同组药品中医嘱名称重复			
药品类医嘱	开始时间	住院药品医嘱成组医嘱开始时间、频次、用法、医嘱状态需要一致,不为空			
医疗费用明细信息	收费类别	住院收费类别必须同时包含(西药、中成药、卫生材料、检验、检查、手术、中成药、床位、幼稚等)			
医疗费用明细信息		成药、床位、护理等) 门诊费用表中数量单位取值有问题,如100盒			
医疗费用明细信息		住院费用表中数量单位取值有问题。如100盒			
院内患者基本信息	性别	男女各不能低于40%			
诊断记录	诊断序号	诊断序号是否为int型,不能为1.0(诊断表)			
修断记录	设施序号	诊断序号是否为int型,不能为1.0 (编案)			
		同一个证据关型、证据状态下,证据名称重复			
设施记录	诊断名称	诊断 8.柜 起代码不能同时为空			

不断增加的质控规则

4000+通用规则

零时项目规则

300+现场定制规则



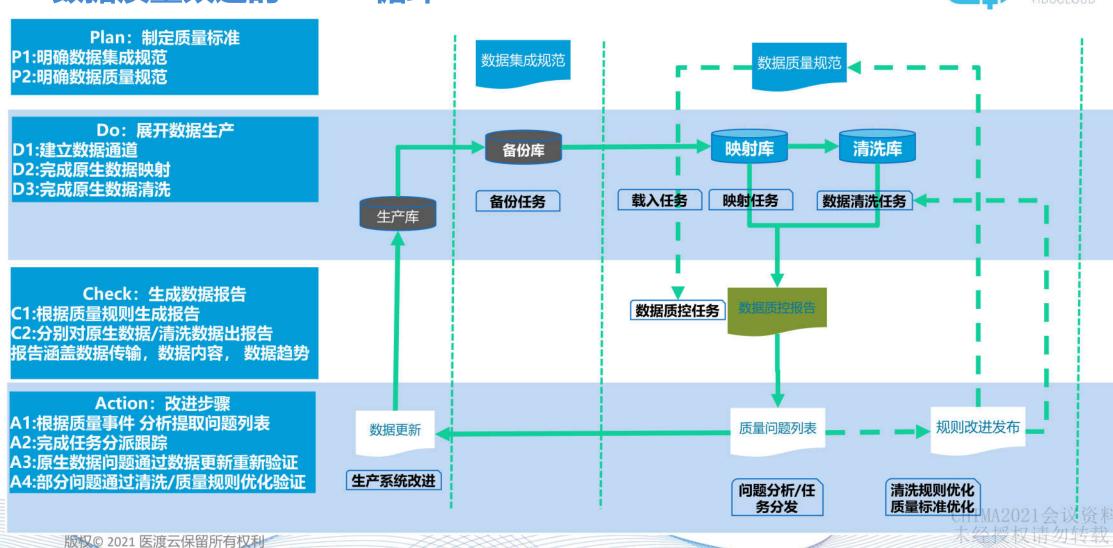
从多维度多指标到单场景评分

		业务指标	流程效率	病种质量
完整性	项目空值	*		
	流程文档		*	*
规范性	专业/病种/手术费用 项目字典	*		
	临床诊断描述		*	*
一致性	时间标记		*	*
stema Metro vision in la est	临床文档描述		*	*
准确性	阈值约束	*		
	基线偏离约束	*	*	*
唯一性	唯一性 患者/住院/医嘱			
关联性	患者/看诊/病历	*		
	医嘱/费用/报告		*	*



数据质量改进的PDCA循环







从多维度多指标到单场景评分: 简化公式

质量评分S=K1*[K2*a+K3×(1-a)]*100

1、K1为定性指标,只有0或1两个分值;

判定条件: 上传数据量<应传数据区间,K1=0,

上传数据量>应传数据区间,k1=1

应传数据区间=本医院前24个月上传数据量平均值±3个标准差。

- 2、K2为定量指标,需根据各医院业务表实际上传字段的填充率统计计算。
- 3、K3为定量指标,需根据各医院实际上传业务表的关联率统计计算,只统计有强业务关联逻辑的表组合,例如发票与费用记录表,费用与就诊记录表等。





从多维度多指标到单场景评分: 总分驱动



国家药品使用监测平台的数据质量控制





挑战

数据维度

药品配备目录 药品出入库数据 药品使用数据

挑战

机构数量

100%三级医院 30%二级医院 10%基层医疗机构 机构总量13000+家





应对

前置质控

利用大数据技术建立YPID 前置质量控制



11136个医疗机构

2441个西药种类

9128个西药规格

3873个中药种类

9294个中药规格



2019 全国公立医

全国公立医疗机构 药品使用监测报告

887299



PDCA

数据提交在线质控数据汇总终末质控





规则为底 场景驱动 化繁为简 闭环改进



