1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3000752/>

CDM模型用于药物安全监控，使用PHARMetrics和GE两个数据库的电子病历数据作为数据集（去隐私化），对罗非昔布和心肌梗塞患者两种疾病药品研究

1）文章的CDM创新模型与OMOP模型相似，包含进行药物安全性分析所需要的基本数据元素，并讲CDM模式描述为实体关系图。

2）文章提出以病人为中心，将病人的医疗经历组织到“Person Timeline”表中，以促成纵向分析。记录病人的基本信息Person（Patient） >>> Person Timeline

3）术语标准化：病症—ICD编码 药物术语—SNOMED-CT、MedDRA

药物数据—字符串归一化工具，来适应临床词汇的需求

condition data—UMLS工具

4）将转换后的数据加载到关系型数据库中（数据的大容量以及对数据访问和查询效率的需求）

5）数据转换过程中每个步骤都进行数据统计，来评估数据转换对数据内容的影响。

在转换过程中，人数保持不变，并且本地数据和转换后的数据相同。例如，患某疾病的人数在源数据库与转换后数据库中比例应该相同。量化，事先统计

6）规则制定需要跟据具体的疾病来具体分析。例如急性疾病可以选取30天连续时间窗来观察，而例如恶性肿瘤等慢性疾病则需要调整。

7）如果源数据库中包含CMD模型不支持的数据，则CMD中不会出现此类数据，会造成数据缺失，质量控制的时候需要注意。例如，文章值未提取实验室观察值，lab项目可能包含与主题相关信息，但是未统计在内。所示设计CDM的时候要考虑可扩展性。

2. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6284277/>

使用OHDSI OMOP CDM进行生存分析：以肝内胆管癌为例，收集分析了ICC患者的HER数据。

1. 采集ICC患者的：人口统计学数据、病史、吸烟状况、实验室结果和肿瘤特征相关数据映射到OHDSI概念。
2. 使用概念图将患者数据加载到CDM中去（考虑OMOP工具）
3. 开发基于OPOM接口，用于生存分析。
4. 将基于CDM的分析结果与传统统计分析结果进行对比---对比实验
5. **问题**：OMOP词汇表对于成像数据、病理数据和模块化数据词汇表示不完整（扩充词汇表）。映射和值规范化不准确造成语义缺失（提前处理，将非结构化数据转化为结构化数据）
6. 缺失值处理：使用R Mice（<https://cran.r-project.org/web/packages/mice/index.html>）来估算缺失值
7. 标准化术语：Athena（<http://athena.ohdsi.org/>），两人审查映射概念并达成共识。
8. R'survival'软件完成生存分析接口。