**数据收集：**

第1章定义了评店患者结局的患者登记的主要特征。包括以统一的方式对数据元素进行明确的、前后一致的定义来收集每一个患者的数据。和随机对照试验一样，病例报告 表是登记数据结构的样板。病例报告表是一个纸质或电子化的数据元素列表。登记数据 库方案中患者水平的变量，代表了病例报告表中的数据元素和数据录人选项。定义登记 的病例报告表和相应的数据库方案是登记开展数据收集的第一步。第 5章中描述了登记 中数据元素的选择。

有两个相关的文件是数据库方案的必备部分：数据字典(包括数据的定义和参数)和 数据验证的规则(也称为查询和输人验证）

数据宇典描述了数据元素，以及如何解释这些数据元素。数据字典包括了登记中使 用的每个变量的详细说明,包括变量的来源、所使用的编码信息和通常取值范围(如果有 的话)。例如,术语“目前吸烟者”应该被定义为是否是“吸烟者”，指吸食烟草或其他物质， “目前”指现在或在最近一段时间内是否吸烟。很多心血管病登记，如“跟着指南走”冠状 动脉疾病计划口,将“目前吸烟者”定义为在过去的一年内吸食烟草的人。 数据验证规则指的是按预先定义的规则在数据进人数据库时进行逻辑检查，无论是 数据范围(如心脏的收缩压应小于 300 mmHg)还是同一个惠者相关数据字段之间的逻畢 一致性;这些内容在数据清洗部分有更详细的介绍。虽然登记的数据库结构和数据库要 求都尚末标准化,临床数据交换标准协会正在积极地致力于用标准化的概念和格式来建 立数据交换和可携带的代表性模板2。第5 章进一步讨论了汶此同时适用于登记和临床 试验的模板。

**程序、人员和数据来源**

数据收集过程需要在规划登记运作时就子以仔细考虑。成功的登记依赖于一个可持续的工作流程模式，与临床实践过程相结合，对医师、护士、药剂师和患者的打扰最小(见第9章)。登记项日会从医疗卫生工作者或可能参与的研究人员最初的投人中获得巨大的效益。

**预实验**

预试验是指在登记完全启动之前从可能的参加者中收集数据。对比第 2 章中讨论过的聚焦登记是否能够实施的可行性试验，预试验关心的是登记应该如何实施。预试验形式多样，可以是检测程序的一个部分、病例报告表或数据收集系统，也可以是在部分患者或医院中实施登记。

有效的预试验的关键是在登记启动之前进行，以便应用预试验的结果对于登记的实施加以调整。通过预试验可以评估登记的可理解性、接受性和可行性，以及在尽量契合患者生活方式和医务人员的日常医疗活动情况下,患者登记过程的影响因素。例如，有些数据来源可能不适用于所有的患者。第5章中对预试验有更加详细的描述。

**过程的记录**

每一个登记的数据收集程序都应该在详细的手册中有明确的定义和描述。术语“手册”指的是任何适当形式的参考资料，包括复印件、电子文档或以互联网或软件为基础的

系统。因为预期目的不同,各种登记的手册细节也不同,但必需的信息大致相同,都包括协议、政策和程序、数据采集工具，以及所有数据元素及其定义的清单。如果登记中有可选项(即不需要每个思者都填满的选项)，则应该予以说明。

除患者的纳人和排除标准外,筛选的过程也应详细说明,所有以医院为单位的筛选活动和任何监控或审计的文件材料也应该保存。对于抽样，则需说明所用的方法或系统,应向各医院提供工具以简化现场的抽样过程。手册应清楚说明患者的编号是如何产生的，以及如何避免重复。也应当对数据收集员必经的培训加以说明。

如果使用纸质的病例报告表,则手册中应当特别说明纸质的病例报告表是如何使用的,以及表格的哪些部分(如二联或三联复制表格)应当保存、复制、提交或存档。如果使用的是电子病例报告表,则应提供清楚的用户使用手册和说明。这些程序对于登记的所有人员(也包括要求登记质量保证的外部审计员)都是重要的资源。

标准化程序的重要性在于保证登记使用统一和系统化的方法，使收集到的数据不会与实际不符。同时，某些程度的定制数据录人方法是必需的或是被许可的，以保证某些医疗实践中的特定医院和患者亚组得以参加。正如在第9章中所讨论的，如果登记为参加的医院提供报酬,那么支付报酬的特定需要应清楚地记录在案,这此信息应该与登记文件一并提供。

所有与数据收集有关的人员都需要经过确认，并描述他们的工作内容，以及在数据收集和处理过程中承担的“角色”。这些“角色”包括患者、医师、数据录人人员、现场主管、咨询台、数据管理人员和监控人员。任何“角色”必要的材料和资格证书都应该在登记文件中加以详细说明。例如，有些登记要求个人文件，如个人简历、签宇的协议、遵守登记程序的承诺书,或完成指定培训的证书。

**数据来源**

登记的数据来源可能包括来自患者的新收集到的数据，由医师报告或来源于病历档案的新的或已有的数据，以区附属的患者信息，如实验室数据。由于用来评估患者预后的登记需要采用统—和系统化的数据收集方法，所有与数据有关的过程包括允许的数据来源、数据元素及其定义、有效性、可靠性，以及其他从各个来源收集的数据的质量要求应该为数据收集人员预先确定并定义好。正如在下面的质量保证部分所讨论的，数据的质量取决于整个数据的收集和处理链。因此,决定登记数据的正确性和质量的根本是链接数据的少而精。

第6 章中,根据数据与登记目的和实施方案之问关系,将数据来源分为主要的和次要的。主要数据来源包含为登记的直接目的收集的数据(即数据主要用于登记）。次要数据来源包括除登记外的其他目的收集的数据(如常规的医疗护理、保险理赔程序)。在下文种将整合和扩展这些定义。

**患者报告的数据**

患者报告的数据 患者报告的数据是按登记目的专门从患者处收集的数据，而不是由医师解释或来自 间接的数据(如实验室的数据、制药公司的记录)。数据范围可能是基本的人口统计学特 征,或者经验证的患者结局报告量表的分数。从运行的角度来看,直接从患者处获取数据 时有许多不同的问题需要考虑，从表达(如字体大小、语言、阅读水平)到技术(如纸笔形式 的问卷调查、计算机输人、电话或语音输人、便携式的患者日记）。这一水平上的错误可无 意中产生患者的选择性偏倚，不真实的结局判断或明显影响成本。病例报告使用特定的 语言或技术，又可能会对参加的对象产生限制。某些特殊疾病的患者可能会很难应用特 定的技术(如宇体过小对于视觉障碍患者、纸和铅笔对于风湿性关节炎患者）。其他的选 择,如提供- 一种病例报告结局的工具，其格式或提交方式都与经验证的工具不同(如使用 调杳问卷而不是面访),则可能导致无效的结果。

**医师报告的数据**

医师报告的数据或从医师处获得的数据也可以分为主要的和次要的。举一个例子，登记可能需要临床医师使用专门的分级量表(如美国卫生研究院的脑卒中分级标准)，而在常规的临床实践中他们却很少用到。有些变量可以直按由医师为登记而收集或从病历档案中获得。必须由医师直接收集的数据元素(如由于—种特殊的疾病定义或评价种特定的并发症所需信息可能并不在病历档案的常规记录中)则需要详细说明。这些分工十分重要,因为它决定了谁能为一个特定的登记收集数据，或是医师必须为登记过程中的患者记录病历档案中的变更内容。此外,登记中产生的错误类型(将在质量保证部分讨论)会因主要的和次要来源数据的使用程度，以及其他因素的不同各异。例如，利用病历摘录进行的登记，可能更容易产生解释性错误！。

**数据摘录**

数据摘录是临床医师之外的数据收集员从患者的临床报告提取数据的过程。虽然体格检查，如身高、体重或实验室检查，如白细胞计数等是直截了当的，但一般来说摘录涉及不同程度的判断和解释。

在数据摘录时，清晰的描述和标准化的定义对于保证数据质量和防止解释错误至关重要。资深的登记人员应当成为现场数据采集人员,设置和落实工作程序，以保证现场的数据采集人员能够与指定的登记人员保持密切联系,咨询有关特定的定义和临床状况等问题。那此时间跨度很长的登记项目，如用作监测的登记，需要-一种运行机制，阶段性地对数据定义进行审核，确保数据元素及其定义的时效性和完整性，以及增补新的数据元素及其定义。通常这种更新的原因是登记启动后引人了一个新的产品或治疗方法。

从未被格式化的纸质文件(如医院病历)中摘录数据是一项费力和之味的工作,尤其是其中包括自由文本,通常需要一个阅读人员。而这个阅读人员，他的任职资格可能从受过培训的“医疗记录分析员”或其他卫生专业人员,到未受过培训的助理研究员，可能需要辨认模糊的手写宇体，翻译晦涩的缩写词和首字母缩写词,以及理解临床内容以摘录充分、必需的信息。登记人员应开发出正式的病历摘录指南、文件材料和编码表格，供分析和审核人员使用。

一般来说,指南应包括对纳人登记的特定类型数据(如特定的诊断和实验室结果)的搜寻说明。分析员经常需要对数据编码，使用代码书(如国际疾病分类第9版的代码)中的标准代码、或使用登记内部特别的代码(如从 1~5 的疾病严重程度)来对应临床病历中相应的文本诊断。

所有数据摘录和编码的说明必须仔细备案，并合并到登记的数据词典中。由于“噪声”存在于未结构化的纸面文件(如伪造的标记和难辨认的宇迹)，以及自然语言的不精确性，由不同摘录者从同一份文件中摘录的临床数据可能存在差别,这是登记中一个潜在的误差来源。

为了减少这种潜在的误差来源，需要对登记的实施方案和程序进行适当的培训,最重要的是数据的定义及其解释。应该对所有登记人员进行培训，对于非临床医师的数据摘录者来说尤为重要。培训时间的长短取决于数据来源的性质(病历或病例报告表)、数据的复杂程度和数据项目数量的多少。各种各样的培训方法，从现场会议到网络会议再到交互式多媒体录音录像，都有成功应用的经验的。培训通常包括使用相同的病历进行测试性摘录。在某此方面,对于某些目的，最佳做法是培训数据摘录员使用标准化摘录表格。这种标准化测试既可以进一步用于获取病例报告表、定义、编码说明的摘录员之间信度的数据，又可以确定某个数据摘录员是否能够达到登记预先设定的最低标准。依赖于从医疗病历中提取数据的登记,应当考虑对病历摘录相关的实施特性加以报告,如摘录员之间的信度(0。对医疗病历摘录的标准化必须注意的关键事项是：

标准化的材料

标准化的培训

用标准化表格进行测试

报告摘录员之间的信度

**电子病例（EMR）**

电子病历 (electronic medical record, EMR)是指由授权的医疗卫生机构中的临床医生生或工作人员创建、收集、管理和查阅的个体健康相关信息的电子记录。电子健康档案Celectronic health record, EHR)是由全国性协同工作标准认证的多个医疗卫生组织中获得授权的临床医师或工作人员共同创建和管理的个体健康相关信息的电子记录，比电子病历要复杂得多口。出于讨论的目的，我们仅涉及电子病历十分有限的功能。

电子病历(和电子健康档案)作为登记的临床资料来源将发挥越来越重要的作用。医学界目前正处于一个过渡时期，患者病历的存储正从传统的纸质病历向电子病历转变电子病历的主要功能是将同一个患者的所有临床电子数据集合到一个数据库中，同样的，纸质病历也是整合各个工作人员和诊疗部门的纸质记录。根据应用范围,电子病历可能包括患者的人口统计学资料、诊断、疗程、病程记录、订单、流程单、药物和过敏情况。电子病历的原始数据来源是卫生服务提供机构。数据可以通过手工输人和医师办公室或病床旁的触摸屏录人电子病历。此外,电子病历系统通常与附属系统，如实验室、药房、放射科和病理科有接口(将在下文中讨论)。附属系统通常有自己的数据库，将数据录人自己的数据库，并向电子病历输出相关的患者资料。

由于电子病历包括患者大部分已有的临床资料,因此可以成为登记中患者信息的主要来源。电子病历通常不包括登记特有的(主要来源)数据，需要单独从纸质文件和电子表格中采集。在接下来的几年内,可能会建立起适当的电子病历系统接口，可以应登记需要提供登记所需数据，既可以建在电子病历内部(与登记成为一体）,也可以建在电子数据捕获系统中(与电子病历成为一体)。电子病历已经成为一些登记的次要数据来源，随着电子病历的更多应用，这种做法会越来越多。在这种情况下，数据可能从电子病历中提取出来,转变成登记的格式,进人登记,与其他来源的特有登记数据一起进人登记数据库中。在某种意义上,这类似于医疗记录提取，只不过是发生在电子层面上的。两者之间有两个关键的不同点。首先,数据是一次性从所有数据中“提取”出来的。在这种情况下，“提取”

指的是将电子病历中的数据转移到登记的数据库中所需的定位和其他决策。它并不能消除潜在的解释误差，但可以集中这一过程，使规则明确和易于审查,这在本章后面部分的内容还会提到。其次,数据都是电子上传的，消除了重复的数据录人以及与重复录人相关的潜在误差和繁重的劳动力成本。

当使用电子病历作为登记的数据来源时，登记需要的信息如果采用自由文本形式,而不是存储在电子病历中编码过的或结构化的数据，就会发生严重的问题。结构化数据的例子包括按国际疾病分类第9版分类的诊断结果和实验室结果。相比之下,医师的病程记录、问诊记录和放射报告等通常都是口述并以自由文本的形式记录的。电子病历中自由文本格式的数据提取可以由一个医疗记录分析员来完成，随着电子病历的应用增加,从自由文本中提取数据的自动化方法已经发展起来。这一技术被称为“自然语言处理(natural language processing, NLP)”。它使计算机能够从自然语言中处理和提取数据信息。自然语言处理的目标是根据一系列规则和词汇表，使软件能够识别关键词，从自由文本解析出有意义的成分，理解语法结构，解决文字含糊问题。这些成分可以与从电子病历中提取出的结构化数据一起提取出来，提交给登记，且两者都以结构化的数据形式存储在登记的数据库中。

可供使用的自然语言处理软件包已经越来越多(如国家癌症中心的 caTIES 81, 2b2床旁的集成生物学信息四，以及若干商业产品)。然而，自然语言处理仍处于早期发展阶段，尚不能应用于所有目的的病历数据提取。总之,自然语言处理软件已在某此特定的临床领域(如病理科、放射科)运行，其专业词汇已经包括在自然处理软件的数据库中。自然语言处理在各种临床环境中，成功地应用于从自由文本提取诊断信息和药品名称。

随着 2009 年美国经济复苏与再投资法案中健康信息技术相关条款的实施,电子病历或电子健康档案的使用预计会有显著的增长。目前，在数据元茶水平检索的系统中只存储了全美一小部分患者的数据。此外，这些系统中只有很小一都分采用数据元素标准化数据定义、以结构化的形式存储数据，数据元素在各用户之间是通用的。目前临床和科研系统之间的格式的交互[如从第？级健康水平(FL-7)a 到临床数据交换标准协会1的模型〕正引起大量关注。数据的语法和语义的问题也受到了广泛关注《见第5章)。对于未来电子健康档案和登记资料之间的交互使用,采用共同的数据库结构和开放性的互通标准将是至关重要的。第11 章将对这一问题进行深人讨论。

**其他数据来源**

一些采用人群登记的临床资料可能来源于数据仓库,而不是电子病历。其他数据来源的例子包括计费系统、实验室数据库和其他的登记。第。 章详细讨论了其他数据来源的潜在应用。

**数据录入系统**

一旦登记识别了主要的和任何次要的数据来源,登记团队就可以确定数据如何录人登记的数据库中。很多方法和技术都可以用来将数据录人或迁移进人登记的数据库中。包括纸质的病例报告表、直接数据录人、传真和扫描系统、互动式语音应答系统(interactive voice response systems, IVRS)和电子病例报告表。根据这些数据以何种速度进人到达中央资源库,并进行数据清洗、审核、监测和报告，有很多不同的模式。每种方法都有自己的优点和缺陷,而登记要在灵活性(可供选择的方式的数量）、数据的可用性(何时落户于中央资源库)、有效性(所有方法是否都能同样产生清洁的数据)和成本之间达到平衡。合适的决定取决于很多因素，包括数据元素的数量、现场数量、地点(不同国家有不同的本地偏好、语言差异、不同技术的可用性也不同)、登记持续时间、随访频率和可利用的资源。

**纸质的病例报告表**

通过使用纸质的病例报告表,临床医师在临床出诊时记录下纸质形式的临床资料，或其他的数据收集员在临床出诊之后从医疗数据中提取出数据。病例报告表包括从不同来源(如医疗记录、实验室、药房)和多次临床出诊收集的每位患者多种类型的临床数据。在格式化的纸质形式的数据录人数据库之前,需要检查表格的完整性、准确性和有效性。纸质的病例报告表可以直接数据录人或通过扫描系统的计算机化数据录人方式进人数据库。

直接数据录人需使用键盘来录人数据。键盘录人的可变误差率取决于不同的个人，因此通常需要估计错误率,尤其是进行大量的数据录人时。双轨录人是通过比较两个不同的数据录人员之间的差异来计算错误率,进而提高手工录人数据准确性的一种方法。由两个人平行录人数据，且由第三个人来进行审查和误差管理可以进一步提高数据的准确性。直接录人数据的前期数据有效性检查可以显著降低数据录人错误发生的可能性。

因此，选择单人还是双人录人数据取决于登记的最大允许错误率，以及每种方法在特定情况下关键测量指标符合这个错误率的能力。双轨录人数据虽然是登记试验的标准实践方法，却明显地增加了成本。是否需要使用双轨录人，取决于关键测量指标的错误率的降低程度以及双轨录人与其他方法相比能够减少的错误可能性。在某些情况下，通过抽取数据样本进行重新录人来评估数据录人的错误率就可以满足报告的目的。

对于纸质表格的数据录人，可使用扫描仪和特殊软件从扫描图像中提取数据。如果数据记录的是表格中多选框的选中标记,扫描软件能够使用户将每个多选框的定位与文本项目代表的变量值联系起来，并确定多选框是否做过标记。多选框中出现的标记将由软件转变成相应的值,然后传送到数据库中进行存储。如果多选框包括了手写的或打印的文本和数字,光学宇符识别软件一般能从扫描图片中提取数据,然而页面上假的标记往往会引起误差,为避免转换误差,对印刷质量的要求很高。基于自动化系统参数所指定的运算符进行错误检查可用于异常处理。上文中对评估错误率所做的评论也适用于扫描系统。

**电子病例报告表**

电子病例报告表是为了向每个临床试验项目的主办者报告的可稽核的电子表，是按照临床试验方案要求为记录信息而设计的。电子病例报告表可以使数据收集人员(医师或其他的数据收集人员)将医师的报告资料直接录人电子系统中。很多登记的现场人员通常先完成一份代表病例报告表的过渡的纸质工作表，随后将数据录人电子病例报告表中。虽然这一方法改进了工作效率,降低了错误率,但要让所有的电子数据录人都在患者床边、门诊或在忙碌的诊疗活动期间完成仍不太现实。

电子病例报告表起初可能是个本地系统(包括个人电脑、局域网服务器或便携式设备)或直接通过因特网连接或专用网络连接来源于中心数据库服务器。对于有超过一个数据收集点的登记，现场系统收集的数据必须随后与中央数据系统交换。！电子病例报告表可以是视觉的(如电脑屏幕)或听觉的(如互动式语音应答系统等电话数据录人）。不同的方法适用于不同的情况。例如，在一个与案例 21 相似的氣氮平患者登记中,药剂师和医师都可以通过一个电话的互动式语音应答系统或网络系统进行数据录人。这种情况下这一方式的成功是因为药房里电话无所不在,而且电子病例报告表非常简单。

电子数据录人的一个常用方法是使用基于网络的数据录人表格。患者、卫生服务机构、。调杳员都可以采用这一形式进人当地的数据仓库。表格存储在服务器上,可能是位于登记的网站或互联网的任何地方。为了获得数据录人表格，有互联网连接的远程计算机用户需要打开一个浏览器窗口，并进人网络服务器的地址。通常情况下会显示出一个登录窗口，用户使用一个由网站或网络存储器负责人员提供的用户标识和密码进人。一旦用户获得了服务器的认证，就会显示出数据录人表格，这时用户就可以开始录人数据了。正如在数据清洗部分提及的，很多电子系统在数据录人的同时可以进行数据有效性的校验和审核。当数据录人完成时，用户要提交录人表格,录人表格将被发送到因特网的网页服务器上。

便携式的设备如个人数宇助理(personal digital assistant, PDA)和手机也能以网络为基础或以其他形式向服务器提交数据。移动网络最近己成为收集临床资料的一个重要的工具。使无线的 PDA 和手机也能够通过因特网向固定地点的服务器传输数据的软件也已开发。随着无线技术的特续发展和数据传输速率的增加，它们将成为更加重要的患者和医师的数据录人设备。

**数据收集技术的优缺点**

医疗记录或附属信息存储在电子表格中时，它们可以由数据采集员摘录到电子病历中,或在有些情况下自动上传到登记的数据库。登记系统和医疗附属信息的接口设计决定了登记系统从电子医疗信息系统中摘录出可用数据的难易程度，以及电子病历和医疗附属系统的软件提供登记所需数据的能力。然而，随着系统的供应商越来越多地采用新的标准和提高互通性，将数据从一个系统转移到另一个系统变得更加容易了。很多组织正在积极致力于改进标准，包括第7级健康组织10]、国家电子卫生保健合作组织13、美国国家标准技术研究所(National Institute of Standards and Technology， NIST)04等。第第 11章会讨论针对电子健康档案系统的标准和认证。

将数据从一台电脑转移到另一台电脑需要数据接口。如果临床数据以电子病例报告表的形式进人一个当地的储存库或电子健康档案中,数据必须要从当地服务器的源数据库中提取出来,转换成登记所要求的标准格式,并存人登记的数据库以长期储存。这一过程被称为“提取、转换、存人”（extract, transform and load, ETL)过程。除非是当地的数据储存库在变量名及其数值方面的设计与登记的数据库一致，否则数据的对应和转换会是一个很复杂的工程。在某此情况下，数据的人工转换会比开发相应的数据接口更加高效和节约时间。新兴的开放性标准使数据能够直接从电子健康档案转换到登记中去。这一话题将在第 11章中详细讨论。

即使已经开发出当地的电子系统和登记的电子系统之间的接口，也仍然需要考虑与临床附属系统就一条患者信息的检索和转换的标准。典型的是，临床附属资料是储存在一个相关数据库中,这一系统需要运行一个结构化查询语言来查询数据库，以检索特定的信息。结构化查询语言可以通过一个标识符(如医疗记录号码)或具体变量的值或范围(如所有糖化血红蛋白超过 800mg/dl 的患者)来识别个体患者。查询的结果通常以能够通过接口转换和转移到登记系统的文件形式(如 XML、CSV、CDISC ODM)存储。有很多种接口协议可以用于转换数据。

由于数据的定义和格式的标准通常全国不统一，数据从电子病历或辅助系统中转移到登记的数据库时经常发生错误。对数据转换说明书进行仔细评估来解释和搜寻这些错误，是登记协调中心需要确认的关键步骤。进一步地,进行一系列的传输试验和验证并记录下来。最后,由于实际操作中可能产生未测试出的新的格式或其他误差，必领从登记本身中识别和区分出来,因此数据审核必须包括在转换程序中。尽管每个部分的数据都可能被精确转换，但在不同的系统中数据的表现形式不同(数值的差异如“0”和“1”代表的意义、定点数与浮点数、日期格式、整数长度和缺失值)。总之,从电子病历记录中提取数据存人登记数据库的任何系统都应经过认证，且应该定期对数据转换进行抽样，以保证信息的上传在时间上的一致性。

如对临床附属系统的数据审核发现错误数据，即使数据已经转人登记中，也必须通知登记系统。登记软件必领能够收到这个通知,将错误的记录值标上无效的标记，并向数据库中输人新的、改正过的记录值。最后需要确认的是，电子

一电子交换的功能不仅是测试，还句括确认转换数据的完整性和质量。目前几乎没有临床附属系统和电子病历系统具有确定的标准。如果登记的目的是向 FDA 或其他有类似要求(包括电子签名、检查系统、严谨的系统认证»的主办机构或数据接收方报告数据，登记与其他系统的交互方式必须要仔细考虑。

**数据清洗**

数据清洗指的是对数据问题的改正和改善，包括缺失值、错误值、超出正常范围的数值，或与数据库中其他数据有逻辑矛盾的数据。虽然所有的登记都致力于“数据清洗”，但事实上,这只是一个相对的词汇。数据清洗的方式和清洗到什么程度必须要在数据管理手册中预先说明，明确需要清洗的数据元素，描述数据有效性的规则,或对超出正常范围数值的逻辑检查,解释缺失值是如何产生的，以及处理与数据库中其他数据有逻辑矛盾的数据。

**数据管理手册**

数据管理员应该开发出供评审员和数据录人人员使用的正式的数据审核指南。指南中应该包括如何处理缺失值、无效录人（如单项选择的区域内出现了多个选择、在数字的区域内出现了字符数据)、错误录人(如某个性别的患者错误地回答了与另一个性别有关的问题)、逻辑不一致的数据(如同一个患者对两个问题的回答是互相矛盾的）的信息。指南中还应该包括修复这些数据问题的程序。例如，调查表中出现了错误的数据,则需要询问调杳员或患者，或参考可能解决问题的其他数据来源。任何数据审查活动，包括日期、时间和询问的结果，都需要保留文件材料。

**自动的数据清洗**

理想的自动化数据检查是在数据录人之前就存在于数据库中的预编程序。在数据收集现场，使用患者或病历档案进行数据清洗，数据检查时尤其有用。甚至只是相关的、简单的检查修改，如实验室数据的取值范围，都可 能对提高数据质量有明显的效果。很多系统会允许使用更加复杂的数据修改检查，且这些检查确实减少了随后的人工数据清理的工作量。使用数据清理规则来限定数据取值范围的方法有很多，用来预防错误的录人这些方法也可以结合起来使用。对于纸面数据的录人方法,纸质病例报告表的输人过程无法进行自动化的数据检查,但可以在随后数据进人数据库时加以采用。

**人工数据清洗**

在出现意料之外的错误时,数据管理员以人工的数据检杳和“疑问查询”来分析数据。这是一种不需要回到数据收集现场的清洗数据的标准方法(如纸质的病例报告表通过录人或扫描进人数据库)。通过使用算法和手工核杳进行数据摘录和分析来仔细检查数据，数据管理员识别出错误，并将生成的“疑问查询”发往现场。即使以电子病例报告表为基础的数据录人包含数据验证规则,也可能无法完全确保基于某种目的数据是干净的。在开发数据管理手册和修改检查时，预计所有潜在的数据错误是非常困难的。因此，即使用了自动化数据有效性审查，仍需要进行一些人工数据清洗。

**疑问报告**

根据数据管理手册,或者数据管理员因为某此目的进行的同期审核,登记协调中心需要定期对收到的数据质量生成疑问报告。这此报告的内容各有不同,取决于为达到登记目的需要进行什么样的数据清洗，以及已经进行了多少自动数据清洗。疑问报告可能包括缺失值、超过正常范围的数据和逻辑不一致的数据(如男性患者的病例报告表中出现妊娠试验阳性的数据）。也可能识别出数据的昇常趋势，如与患者的历史均值或临床上已确定的正常范围相比,实验室检验结果突然上升或下降。合格的登记人员有责任与指定的现场人员一同检查数据的异常趋势。现场最有效的方法是就疑问目的或登记人员所虑提供—名联系代表。根据现场的数据记录和数据来源的可用性来审查和对“疑问”作出反应,有时要解决所有的疑问很困难。推荐建立一个系统化的方法来最大限度提高现场的应答率。

**数据追踪**

对于大多数登记的目的，对数据接收(纸质的病例报告表）、数据输人、数据清洗和其他参数的跟踪是主动登记管理的一个重要组成部分。通过比较指示性指标，如与实际相对应的预期患者参加率、病例报告表完整性、疑问率等,登记协调中心可以发现问题并进行纠正无论是在个体水平上还是在整个登记水平上。

**数据编码**

正如第 5 章中进一步讨论的,标准化代码字典是用来将登记数据和其他数据库合并的越来越重要的工具。如同“健康信息团体”采用的标准，除非有特殊原因,这些代码标准应为登记常规应用。虽然这些代码应该在登记计划阶段就包括在数据字典中，但接口中不可能包含所有代码。-一些自由文本可以作为结果进人。当自由文本进人登记时，使用标准化宇典对这些数据进行编码(医药监管术语辞典、世界卫生组织药物命名、医学分类命名法)的方法是合算的。重新编码需要花费成本，一般来说,编码的数据元素应该仅限用于分析或与其他数据库合并所需的数据元素，如用于为多个登记和研究而运行的共同安全数据库。

**数据存储和数据安全**

当表格上的数据进人登记的电子数据库之后,表格本身连同数据录人的记录,都要按档案管理规定定期保存。因为数据错误可能会在数据存储于登记数据库之后很久才发现。错误可能发生在患者和调杳员填写 原始调查表时，也可能存在于数据录人过程中。检杳原始调查表和数据录人记录就可以发现错误的来源。如果错误在原始调查表上，则需要重新调查患者才能修改。如果错误发生在数据录人时，则应更改录人的数据并更新数据库。那时，错误的登记数据可能已经用于生成报告或建立人群研究队列。因此，不能仅仅用正确的数据来代替错误的数据，登记系统应该有能力将数据标记为错误数据，而不仅仅是删除它,并插人正确的数据为将来所用。

一旦数据录人登记，登记必须定期对数据进行备份。通常有两种基本的备份方法，登记协调中心应考虑将两种方法都列为最佳做法。第一种方法是实时的磁盘备份，通常是由登记服务器使用的磁盘存储器完成的。第二种方法是使用可移动介质(如磁带、只读光盘驱动器、数宇化视频光盘)进行常规(如每天的）备份。在第一种方法中,数据存储在登记服务器的磁盘上，被自动复制到两个或更多的硬盘驱动器上。以最简单的“镜像”为例，登记的数据被储存在一个初始的磁盘上,一个额外的复制品就被储存在镜像的磁盘上。如果任何一个磁盘损坏了,数据仍能继续保存在镜像的磁盘上，直到损坏的磁盘被重新替换为止。这一故障用户是可以完全看到的，在故障期间他仍可以向登记的数据库中继续录人和检索数据。也有配置更加复杂的磁盘备份，其中一系列的磁盘都是用于为某个损坏的磁盘提供保护。

灾难性恢复则通常需要用第二种定期修复的方法。理想地，存储登记数据库的日常备份应该储存在可移动的介质里并异地保存。如果发生了登记服务器损坏或登记中心关闭的灾难，备份文件可以放到可运行的登记服务器中,重新存人登记数据库,可能只损失了在两次常规备份间隔内的数据。丟失的数据通常可以从现场的数据储存库中重新装载,或由纸质表格重新输人。其他先进和普及的数据库解决方案和灾难恢复技术可以支持个位手远程数据中心的“备用的”数据库，万一原始的数据库损坏了，仍可以使用备用的数据库,缩短故障时间和防止数据丟失。

**管理变更**

正如所有登记的其他程序一样，变更管理的程度取决于收集到的数据类型、数据来源和登记的整个时间表。在登记实施期间，变更的需求来自两个主要驱动因素：内部驱使的变更以改善登记或提高收集到的数据质量，外部驱使的变更是由登记实施的外环境改所致。

内部驱使的改变通常集中在由现场反馈、疑问查询、疑问查询指向某个问题的倾向、定义、病例报告表的项目设计太差或缺失等引起的数据元素或数据有效性参数需要改变。如果情况的确如此,那么登记可以按照调查现场或数据管理员反馈的信息,增加、删除或修改数据库、病例报告表、定义或数据管理手册。有时候,实质性的变化太多，比如添加新的表格或改变登记流程，最好对新的情况或结局进行核查。外部驱使的改变通常出现在时间跨度很长的登记中,因为所关注的疾病或研究的产品有了新的信息,或有新疗法或新产品已引人临床实践。登记人员流动是另一种类型的改变，而且缺乏标准化和过程记录的话，可能是非常具有破坏性的。

当登记的病例报告表或基本数据库发生品著变化时，更为广泛的变更就可能发生。因为随疾病和治疔方法进展，,数据收集表格和定义需要更新，因此时间跨度很长的登记可能会不时地出现这种问题。处理这种变更的一种方法是锁定先前的数据库，并开发新的数据库。然而更多的时候，是希望先前的数据能够用于现场审核或包含在报告中。这就需要数据迁移。

数据迁移包括将数据从一个数据来源转移到另一个改进的或具有不同结构的数据来源中。这一过程包括将数据从先前的数据库中提取出来，并将数据加载到登记的当前版本中。如果这一过程是人工进行的(人工法用子处理数量较少的数据)且两个系统之间有很多实质性的差别(如两个系统中的数据定义和收集方式是否有所不同)，那么其注意事项就与人工的数据录人或数据摘录相似。对于数量很大的数据，数据迁移通常都采用计算机自动的方法。需要创建出详细的数据迁移标准文件,描述源系统中的每一个数据元素和值是如何转变成目标系统中的相关数据元素和值。一般的假设，例如如何处理矛盾的数据，需要在数据迁移计划的开始部分写明并存档，迁移过程中发生的任何问题都会带来新的假设。迁移到目标系统的数据要经过检查，以确保它们与接受系统的数据字典致,符合数据有效性参数。数据一旦迁移，就需要通过一个质量控制过程来保证迁移的准确性。

适当管理变更对于维持登记至关重要。一个贯穿始终的变更管理方法，包括决策、文件材料、数据映射和确认，是维持登记的质量和数据有效性的一个重要方面。某个变更管理过程可能取决于登记的性质和类型,而以评估患者预后为目的的登记，其变更管理至少要求有以下结构和过程：

•程序的详细手册：正如在之前所讨论的,定期更新的详细手册包括所有的登记政策、过程、协议，以及所有数据元素和定义的完全的数据字典对于登记的运行是至关重要的。手册也包括登记中变更的管理和记录的关键因素。

•主管部门：正如第2 章中所讨论的,基于多种原因，登记都需要设置监督和咨询部门。其中最重要的一点是定期变更管理。这个主管部门的基本职责是保持登记手册和数据定义的及时更新。大型的前瞻性登记(如全国外科质量改进计划)经验表明，需要委托一个定义特别委员会来对数据元素和定义进行更新。

•培训准备：如上所述，人事变动是登记中经常会遇到的问题。为了应对任何意外的登记的人员变动或流动,或定期向登记中录人数据的卫生服务机构的变动，应该具备随时提供培训的详细明确的工作程序和设施(见案例 29）。

•沟通方法改变：由于登记经常进行改变,因此在改变即将发生时需要有与各个现场沟通的标准方法和时间安排。

除了搭建以上这此结构,登记还应该从预算的角度(第2章)和分析的角度（第 13章)做好计划应对变更。

**在提供保健服务、协调和质量改善种应用数据**

**提高卫生服务水平**

由于登记越来越多地使用电子表格收集数据，数据收集与提供卫生服务之间的时间间隔大大缩短。这个更短的时间间隔为在患者和人群水平上利用登记功能来改进卫生服务提供了重要的机会。这些功能(表 15)包括产生在个体水平上用以改善卫生服务和协调的输出结果(如决策支持、病例报告、提醒、通知、主动服务对象的名单、教育内容)和提供有助于人群管理、质量改进和质量报告(如风险调整、人群观点、基准剂量、质量报告传输)的工具。一系列的登记最初都是为这个目的设计的(见案例 30）。很多大型的全国性登记15-181 都显示，在医院和私人诊所参加登记期间日常医疗服务有了显著改变。例如，在一个利用从医院比较组织(由医疗保险和医疗补助服务中心开发在线的数据库)荻得的医院数据的一项并行研究中,美国心脏病协会的“跟着指南走冠心病登记项目”参加医

院中的患者,接受了循证提醒和实时的执行测量报告，与没有参加登记的医院中的患者相比,在指南依从性的测量指标上明显更好。

|  |
| --- |
| 输人：获得数据  •识别/招募有代表性的患者(如抽样）  •从多个关键的来源和场所收集数据(卫生服务提供机构、患者、实验室、药房)  •使用统一的数据元素和定义(危险因素、治疗、结局)  •数据的检查和校正(有效性、编码等）  •将不同来源的数据在患者水平上链接起来(患者身份标志的管理)  •保证安全和隐私(如访问权限控制、审查性跟踪） |
| 输出：保健的提供和协调。  •提供决策支持的实时反馈(证据/指南)  •生成患者水平的报告和提醒(纵向报告、医疗差距、概述列表/计划、健康状况》  •向患者和卫生服务提供机构发送有关的通知(医疗差距、预防报告、自我管理）  • 与患者和其他的卫生服务提供机构分享信息  •列出需主动服务的患者/亚组名单  •与相关的患者教育链接起来 |
| 输出：人群测量和质量改进  •提供人群水平的报告  -实时、快速的周期  -风险调整  -包括标准化措施  -包括基准剂量  -向不同水平的用户提供不同的报告  •为探索性研究提供特别的报告  •为人群和亚组的管理提供工具  •产生状态控制信息以促进行动  •协助第三方进行质量报告(传输） |

**特例：与临床表现相链接的产品使用系统**

与临床表现相链接的产品使用系统也称限制使用或有限使用 系统 (performancelinked access system,PLAS)，是登记不以观察为目标的另一种应用。与疾病和暴鳐登记不同的是，与临床 表现相链接的产品使用系统是由主办方作为按承诺开发的，在产晶准人市场以后，作为详细的风险最小化行动计划的一部分，用以加强产品的风险和利益平衡。与临床表现相链接的产品使用系统的目的是通过确保产品使用过程与具体的临床表现测量指标相链接，用以减轻已知的与药物有关的风险。具体的实例包括监测实验室数据的系统，如在患者服用氯氮平期间通过监测白细胞计数来防止严重的白血病，或在服用反应停期间通过常规的妊娠试验来防止这种已知的致畸物的宫内暴露。更多与临床表现相链接的产品使用系统的信息请参见《风险最小化行动计划的发展和应用指南》一书。

**质量保证**

在确定一项登记对于决策的作用时，了解获得数据的过程和数据库中存储数据的质量是至关重要的。能够充分满足质量标准的患者登记(在第1 章和第 14 章中讨论的)越来越被重视，它能够产生关于效果、安全性、卫生服务质量证据的重要手段,为了评估登记用于决策的适用性，必须充分理解登记内部的数据质量。登记计划人员必须考虑如何将质量保持在一定水平，以满足预期使用目的(如下所述)，而且还应考虑如何开发出适用于此登记项目的质量保证计划。登记的执行人员必须对这些质量保证活动进行评估和报告。

根据登记的预期使用目的不同，质量保证的方法也有很多种。主要用以提供決策依据的登记2《如覆盖面测定、产品安全评估、按绩效付款)会比只描述疾病自然史的登记需要更高水平的质量保证。质量保证活动通常列为以下三个主要方面：①数据的质量保证;② 登记过程的质量保证;③ 计算机系统的质量保证。由于很多登记是大规模的，质量保证水平可能会受到预算的限制。

为了使足够的质量保证需要与特定目的合理的资源支出保持平衡,强烈推荐采用基手风险的质量保证方法。基于风险的方法集中于误差的最重要来源,或从登记目的角度考察其程序性失误。这样的错误来源可以在登记初始和设计阶段就定义好。如下所述，不同目的的登记可能面临不同错误来源的危险,集中于不同的医疗实践以及评估水平。如果研究结果要在各个登记之间加以合并或比较,对于特定目的采用标准化的方法(如全国性的绩效评估)很可能在未来更加普遍

**数据的质量保证**

结构、过程、政策和程序需要落实到位以确保登记范围内的数据质量，并且确保下列类型的错误得以控制，包括：

•解释或编码错误：此类错误的一个例子是两个不同的数据摘录员从同一个患者的病历记录中提取数据元素,却从同一份表格中提取出了不同的数据。对某种疾病状況或治疗方法的编码错误也属于解释性错误的范畴。避免和发现解释性错误的方法包括对定义进行充分培训、利用标准表格进行测试、摘录员之间一致性的检杳和报告，以及重复摘录。

•数据录人、迁移或转换精度所致错误：当数据未准确地进人登记数据库时，就可能发生这些错误。例如,实验室数据2.0被错误录人为 20。避免和发现这些错误可以通过预先的数据质量检查(如数据的取值范围和有效性检查)、数据样本的重新录人以评估准确性(根据研究目的确定数据抽样的百分比）、严格的数据清洗等方法来实现。

•倾向性的错误：故意歪曲数据(通常被称为数据做假)的一个例子是相对于手术调整风险结局,登记夸大报道了术前患者风险，或只选择结局较好的病例来报道(通常被称为选择偏倚)。要想避免和发现倾向性错误是很困难的。有一些方法包括检查各个数据收集现场之间数据的连线性,将记录日志与其他来源的信息进行核对(如账单数据），以及执行随机的或寻找原因的现场稽杳(包括监测原始数据）。数据质量保证的步骤包括：

•培训：用标准化的方法培训数据收集人员和摘录人员。

•数据完整性：在可能的情况下，即时反馈现场提出的问题，如出现缺失值、超出正常范围值和逻辑不一致的数据。

•数据一致性：在不同时间和不同现场问进行比较。

•抽取—个样本进行现场稽查：对记录日志和程序或样本数据进行稽杳。

•寻找原因的稽查：使用预先确定的和基于大量数据的方法来确定可能产生不准确数据或存在倾向性错误的现场，如招募和记录日志的差异,数据范围过窄,招募率过高或过低。为了进一步减少或识别这些误差和保证数据的整体质量，需要考虑以下几点。

**在每个现场指定一个数据质量保证的负责人员**

向登记提交数据的每个现场至少要有一个数据质量保证的负责人员，无需考虑这个人是否同时参加数据收集。现场的协调人员需要完全了解登记所有的协议、政策、程序和定义。现场的协调人员应该保证所有送往登记协调中心的数据都是有效和准确的。

**对合格人员队伍的培训和维持进行评估**

对合格人员队伍进行周密的培训并记录在案，不论对于现场还是登记人员，都是保证登记质量所必需的。如前所述,对于登记的所有人员进行适当培训,详细和全面的操作手册至关重要。必须常规地对医务人员进行关手患者登记的要求、产品正确使用的认知性测试(调查),用以监督医务人员的基础知识和是否按照患者登记要求执行的情况。当调查结果显示知识维持情况不佳时，就要计划开始再培训。所有的登记培训计划都应提供对数据收集人员关于登记的知识和数据收集能力进行定期评估的方法，尤其是定义或程序发生变更时。

**数据质量审查**

如上所述,登记中数据清洗的水平受到登记目的、收集到的数据类型(如临床数据或经济数据）、数据的来源(主要的或次要的）、登记的时间跨度(历时 3个月的随访或历时10年的随访)的影响。这些登记的特征通常影响到人工或电子数据中产生质疑问题的数量和类型。除了识别缺失值、错误或超出正常范围值、或数据库中存在逻辑不一致的应答结果外,特别培训的登记人员可以审查有疑问的数据，以判断可能的误差趋势，确定是否需要额外的现场培训。例如，这些登记人员可能会发现，不论是一个现场还是整个登记的所有现场，有关某种患者结局的问题或电子病历某一变量产生了大于正常比例的有疑问的数据。利用这些信息,登记人员可以对现场实施定向跟踪，以减少今后某种结局问题和电子病历这一变量的疑问数据率，并就正确的定义进行重新培训。这些类型的“培训技巧”也可以写在登记的通讯中，是一种与登记现场保持频繁而无干扰沟通的方法。

如果登记目的要求对来自各登记者的人库数据进行更加严格的检查，登记的计划人员需要进行现场审计。和上面讨论的问题一样,一个特定登记的审计计划会受到登记目的、收集到的数据类型、数据来源和登记的时间跨度的影响。此外,登记的开发人员必须在稽查范围与对整个登记费用的影响之间找到平衡。以登记目的为基础，登记的开发人员可以定义出需要重点审计的特定数据城(如关键的效果变量和不良事件的数据）。术语“审计”指的是审查和确认，可能是在现场进行(有时称为监测）或在现场外进行，可能是大范围的也可能十分有限。审计可以抽取所有现场的一个随机样本（如 5%~20%的登记现场）;或以寻找原因为目的(意味着仅在可能有问题的现场进行，如其中一个现场与其他大多数现场相比出现明显的异常值);也可以是患者的随机样本;或根据地理方位、医疗服务地点(学术中心或社区医院)、患者的参与率、出现问题的概率(以风险为基础的审计策略)使用不同的抽样方法。

审计数据质量的方法应该能够根据登记日的发现误差最主要的来源。例如，使用绩效测量的登记会有较高的排除高风险患者的风险(选择性偏倚)。而审计的目的除了数据的准确性之外还包括数据的外部来源，以核实记录日志的信息(如账单数据)(见案例32）。最后,登记的时间跨度可以帮助确定审计计划。随访时间短的登记(如随访期为 3个月)可能只需要在研究结束后、数据库关闭和进行数据分析之前进行一轮审计。例如，在 OPTIMIZE- HF 登记中（见案例 26），以预定的标准为基础，在10 000 名患者的记录中随机抽取 5%的样本与原始档案进行数据质量审计12。对于需要长期随访的登记，登记人员需要在登记期间内每 1~2年实施一次现场审计。

除了以上提到的特征之外，有明显人员变动的现场应该考虑首先进行审计，以证实对于新的人员是否进行了足够的培训，并尽快评估人员间一致性。为了尽可能降低对登记的观察性特征的影响，审计计划应当列人登记手册中。

用来评估患者预后和产生科学信息的登记,以及利用病历记录摘录数据的登记，必须根据研究目的采用足够科学严密的方法评估数据收集人员之间的一致性。例如，在一个进行大量数据摘录的登记中，应设计并发布一个细致的系统，评估摘录员之问一致性;除了要求数据摘录人员具备一定程度的熟练水平,还应对一部分表格根据预定标准重新摘录。这些摘录的可靠性指标(如 kappa 统计量）需要反复测量并报告，(现场或远程)稽查之后，需要与现场人员就稽杳结果进行面对面的交流，连同随访结果的书面沟通和改进途径都要记录在案。根据是否符合登记目的的情况，登记的主办者可能要求现场提出改进措施。提高现场的顺应性也可以通过定期与现场就患者登记系统收集到的数据进行交流。

**登录的程序和系统**

**登录程序的外部稽查**

为保证登记特定目的的数据质量水平,登记开发人员一旦确定需要外部稽查，就应该按照与预先制定的标准一致的方法来实施。预先制定的标准中应当包括对患者参与率高的或先前稽查提示存在需要注意的现场、或基于现场的不同经验水平、严重不良事件报告率或所发现的问题进行监视。登记协调中心可对现场的一些样本进行监测，可集中于个或几个地区。外部稽查的过程可以从审查登记程序,采访现场工作人员，到核对记录日志,追踪病例记录个案。

在登记程序的质量保证中，有一个完整和详细的描述登记政策、结构和过程的登记手册是十分重要的，再怎么强调也不为过。这样一本手册既是实施稽查的依据，又是一种记录稽查所带来的改变的方法。数据质量稽查，以及登记过程稽查发现的问题都应当及时反馈,应该与所有的项目相关人员交流，并记录在登记手册中。

**保证系统的完整性和安全性**

数据管理过程的所有方面都应该采用严格的生命周期方法进行系统开发和质量管理。每一个过程都要有严格的定义并记录在案。尽管部分种类可能有所变化,以下所述的概念与很多软件行业标准和医疗卫生行业标准是一致的(如联邦法规第 21 章第11 款，法律安全标准)。登记协调中心所具有的内部质量保证功能应当对这些过程和程序进行常规稽查。在由第三方而不是登记协调中心来执行的与登记系统和数据有关的活动时，则通常要进行风险评估,并且这些活动隶属于登记协调中心的常规稽查。

**系统的开发和确认**

所有用于患者登记的软件系统都要遵守软件开发的一般原则,包括以下软件行业公认的系统开发生命周期模型。与此同时，系统开发的质量保证要使用己核准的说明书为每个项目制定有效性的验证计划。由专业人员开发出作为测试的案例,并系统地执行登记,然后对登记结果进行记录和审查。根据法律法规的要求，最终的验证报告通常是书面和经审批的。悬而未决的产品和过程性问题将继续由问题跟踪系统或纠正措施 /预防措施(CAPA)来进行跟踪。同样的，开发和验证的过程应有文件记录并定期稽查。由适当的小组收集、总结和评估这些审计信息，旨在进一步改善过程和改善质量。

**安全**

所有的登记都保存健康信息，因此安全是一个重要的问题。本部分讨论的是安全管理在美国开展的登记中的应用，在其他国家收集数据的登记需要依照更多的或不同的规定。1996 年的健康保险流通与责任法案的安全规定列出了由健康计划、卫生服务资料交换中心和某此卫生服务机构实施的保护电子健康信息的安全标准2。尽管这些标准是针对电子的需要保护的健康信息，但这些原则本身还可以用到更广泛的范围。确保安全并不只是简单地依靠技术，还需要清晰的过程和程序。要按惯例分派好全部的安全责任。安全程序通常需要很好地记录在家和公示。这些文件材料也可以用于培训工作人员。有些登记还可能保存有个人信息，如联系患者和提醒他们收集和提交“患者结局报告”所需信息。联邦政府和所有的州政府一样，制定了有关妥善保管个人信息以及要求对个人信息的安全漏洞进行报告的法律。具体的要求各个州之间有所不同。

**系统的安全计划**

系统安全计划由政策文件和精确解释系统规则的标准操作程序组成，包括适用的管理性程序、物理防护、技术安全服务、技术安全机制、电子签名和审计跟踪。这些规则界定了角色和责任。规则包括个体的行为责任界定的政策、基于最少特权原则的访问权限、责任分散的需求。这些原则和伴随的安全行为为登记数据的保密性和完整性奠定基础。这些规则也详细描述了不遵守规则的相应后果。

**教育和培训**

登记协调中心的所有工作人员需要定期接受关于整个系统、安全要求，以及各个患者登记的特殊要求等方面的培训。每个工作人员必须按受与其工作责任相关的培训，工作人员是否接受了适当的培训应记录在案。

**访问权限**

系统和数据的访问权限应以最少特权原则和责任分散原则为基础。不允许任何工作人员有超过工作需要的访问权限，也不允许任何工作人员可以拥有能够避开系统控制和拒绝行为的访问权限。在任何情况 下，只有被授权的工作人员才有访问权限。

**访问权限控制**

访问权限控制的基本作用是对关键的系统和资料进行经身份认证和符合常规的访问。由于储存在电子系统中数据的真实性、完整性、可审计性有赖于精确的个人身份认证,电子签名管理(将在下文中讨论)是一个重要的话题。

对系统和电 子化数据的访问必领以一种途径加以控制，即只有授权的工作人员才能获得系统的访问权。这通常通过一个独特的访问码来实现，比如使用一个独一无二的用户名和密码的组合，这—组合同时分配给每个身份经确认、其工作责任需要访问数据库的工作人员。系统应要求用户定期更改密码，并能够发现可能存在的未授权使用的行为，如多次登录失败,并在此情况出现时让该账户自动失效。身份编码也可以加密成为电子证书,存储在密码保护设备上，或设计成只供用户本人进人的生物识别系统。

必领建立资格认证受损情况下的进人规则。新的密码信息必须通过—个安全的方法发送给个人。

数据收集现场可以连接到互联网，但需要安装人侵检测和防火墙系统,用足够的保护和规则来控制授权用户的访问。桌面系统也应该配有反病毒软件，服务器需要运行最新的安全补丁。在整个登记全过程，都需要检查系统安全，以保证管理、运行、人员和技术控制的正常运转。

**数据飞地**

随着临床资料增多,各种团体和研究人员对临床资料需求总量的增长,新的数据访问方法正在出现。数据飞地是-一种安全的远程访问系统，允许研究人员之间以受控和安全的方式来分享调查对象信息。针对各种检查者，数据飞地使用不同水平的统计性、技术性和操作性的控制。这对于增强数据保护，以及使某一组织能够获得其所需求的数据都十分有用。数据飞地也可以用于允许其他研究者在受控制的前提下获得登记的数据随着登记及其众多的利益相关者对数据使用的增长，数据 飞地变得越来越重要。

**电子签名**

电子签名为个人承担责任提供了一个基础，与稳妥的、计算机化的、具有时间标记的审计性跟踪联合应用时，有助于确保获得一个精确的变更历史。绝大多数系统都使用电于签名。对于向 FDA 报告数据的登记，这样的签名必须符合联邦法规第 21 章第 11 款关于一般签名的组成、使用和控制的标准(11.100，11.200 和11. 300）。然而，即使是那些没有这些要求的登记也应该将这些视为合理的标准。在给一个人分配电子签名之前,核实这个人的身份并告知电子签名的意义非常重要。在电子签名由用户名和密码组成的情况下,应采用管理性和技术性的方法来保证签名的唯一性以及符合密码创建规则。密码长度、字符位置、有效性周期都应基于已有的最佳实例和美国国家标准技术研究所出版的指南。用在电子签名中的密码需要遵守与系统准人控制规定一致的安全和有效期的限制。

**验证**

那些存储符合 FDA 要求的电子记录(或依赖于这些记录的电子或手写签名)的系统都必领按照 1997 年3月20日发布的联邦法规第21 章第 11 款最终规定进行验证。这一规则描述了按管理条例(通常称为“断定规则”)对电子系统和任何提交给 FDA 的电子记录的要求和控制。FDA经常出版非约束性的指南性文件，概述了目前对这些法规的适用范围和应用的解释。目前的指南性文件是2003年8月出版的《行业指南,第11 章,电子档案;电子签名范围和应用》。另外关于电子系统验证要求的文件是2007年 5月出版的《行业指南，临床调查中应用的计算机系统 》，以及2002年1月11日出版的《软件验证的一般原则;对行业和FDA 工作人员的最终指导》。

**资源考虑**

根据登记的总体目标,登记费用的变动会十分大。登记的花费也与登记点总数、患者总数、登记的各个现场的交通便利程度相关。本章中描述的每一个元素都对应着一个相关的花费。表16列举了登记协调中心的一些活动清单。并不是所有的登记都需要或能够负担所有在本章中提到的功能、选择或质量保证技术。登记的开发人员必须杈衡利益和可获得的资源,以决定最恰当的方法来实现目标.

表16 登记协调中执行的数据活动

|  |
| --- |
| 数据管理  •定义所有正在进行的数据质量控制的步骤、程序和测量指标  •定义针对数据所执行的修改检查的类型  •定义电子文件必需的文件格式的详细说明书，以及进度表和转换数据的程序  •定义电子数据的质量标准，以及处理异常值的程序  •开发数据输人的指南  •识别电子审核无效而需人工审核的数据  •编写和维护审核、编码、报告不良事件数据的程序  • 编写和维护归档的程序  • 编写和维护变更管理的程序  •编写和维护能够有效识别突出的疑问、每个现场的疑问类型等疑问跟踪和创建标准报告的  程序  •将疑问数据与过程改进所需要的程序和活动联系起来(如病例报告表的设计）  •通过执行精确的数据库更新,跟踪数据清洗中发现的有疑问的值和错误  •定义登记专用的字典和代码清单  •合适的话，执行数据库南计  •按照书面的详细说明书执行系统和应用程序的用户测试  • 建立质量准则和最大容许错误率  • 评估需稽查的数据点，为稽查找出潜在的数据误差来源  •确定错误的根本原因以提出对过程和技术改进的建议，确保错误不会再次发生(持续改进〉  •确保抽样稽查技术是有效的，并能够为数据决策提供支持  •概括所有其他的数据流,包括外部的数据来源 |
| 文件档系  •进行培训的过程、程序、标准、内容清单,记录在案  •发现数据存在疑问的信号或趋势的程序和标准，记录在案并保存  •对数据库的质量保证活动，记录在案 |
| 报告  •生成病例数据库中缺失值的标准报告  •制造工具以追踪病例报告表并编制病例报告表的目录清单，并报告预期的和实际的病例报表数量 |