arXiv:2611.33074v1 [cs.DB] 24 Nov 2026

从患者旅程创建以对象为中心的事件日志研究论文

Saimir Bala[1], Neven Borak[2⋆]、Maximilian König[1], Anjo Seidel[1],Zeinab Soleimani[1], Lorenzo Famiglini[2], Lev Petrov[2], Jonas Witt[2], 和 Mathias Weske[1]

[1] Hasso Plattner研究所、波茨坦大学、德国
firstname.lastname@hpi.de [2] mama health，德国，
https://www.mamahealth.com
firstname.lastname@mamahealth.io

摘要。 患者与卫生系统之间的互动，以及个体在疾病期间经历的疾病状况，都体现在患者旅程中。因此，患者旅程是研究医疗照护如何提供，以及如何设计卫生系统以更好地满足患者需求的重要资源。尽管从流程挖掘视角出发、基于传统案例概念研究患者旅程的工作已经初步展开，但对象中心流程挖掘技术尚未得到应用。鉴于对象及其关系的丰富结构，对象中心流程挖掘方法在这一语境下尤其具有前景，这一缺憾也因此令人遗憾。本文通过研究如何从患者旅程生成对象中心事件日志（OCEL），弥补了这一空白。基于一系列挑战和要求，本文提出了一种从患者旅程数据创建以对象为中心的事件日志的方法。该方法在我们的行业合作伙伴提供的合成患者队列上进行了评估。

关键词： 对象中心事件数据 · 流程挖掘 · 医疗保健 · 患者旅程

1 引言

将流程挖掘应用于医疗保健领域，可以带来前所未有的洞见，并有助于提升医疗照护质量以及为支持医疗照护而设计的系统 [1,9]。通常，流程挖掘会从医疗服务提供者的视角应用于医疗保健数据。这一视角有助于识别医疗保健系统中的瓶颈和低效之处，并改进其流程。然而，这种做法隐含地将患者视为流程的“客户”，从而在三个方面限制了流程改进的范围。首先，它忽视了患者自身的知识，尽管患者往往对自己的临床病史了如指掌。其次，由于数据可用性方面的限制，分析通常局限于

⋆ 第一作者和第二作者对本文贡献相同

单个信息系统（例如单家医院的信息系统），而患者护理涉及多个系统、临床医生和治疗。第三，这些分析所得的洞见往往旨在改进医疗服务提供者的流程（例如医院的流程）。当然，此类改进可能使患者受益。然而，衡量医疗照护质量的最终标准，是患者在整个旅程 [10],中的亲身体验；仅分析信息系统日志中的数据无法完整捕捉这种体验，因此必须直接向患者了解。不过，患者的叙述并不完整，也可能发生变化；它们以自然语言表达，且涉及疾病状况、治疗和利益相关者构成的复杂领域。这三个挑战使得在应用分析方法之前，难以对数据进行建模。此外，患者旅程会同时涉及多个实体（例如患者、疾病状况、治疗）。以对象为中心的事件日志（OCEL） [3] 最适合同时分析所有这些视角。

本文指出了从患者旅程创建以对象为中心的事件日志所面临的挑战和要求。为应对这些问题，我们提出了一种分步方法，涵盖从数据收集到生成 OCEL 的全过程。我们的方法具有若干关键特点，包括对从患者叙述中提取的概念进行时间与语义解析。我们的方法还包含一个映射方案，以指导 OCEL 的创建。该映射方案基于通用原则，可应用于任何患者旅程，将其转换为 OCEL。我们在一个合成患者队列上评估了该方法。

本文其余部分的结构如下。第2节介绍患者旅程及其与流程的关系，随后通过一个示例说明患者旅程。第3节阐述将患者旅程作为流程进行分析时面临的挑战和要求，并提出一种解决 OCEL 创建问题的方法。第4节对该方法进行评估。第5节讨论相关工作。第6节总结主要发现。

2 背景与运行示例

2.1 患者旅程

从患者的视角来看，医疗照护是与医疗服务提供者之间的一种关系，例如诊所、医生或医院。它通过与不同医生、地点和治疗方案的一系列就诊接触展开，这些就诊接触通过交接 [16]相互连接。按计划实施的医疗照护流程称为路径。因此，临床路径指的是计划中的临床干预和治疗方案，以及促成并支持这些干预和治疗方案的活动。当临床路径在实践中付诸实施时，我们便得到一段患者旅程。患者旅程被定义为实际采取的措施以及实际发生的、影响患者医疗状况、功能状态、体验和情绪的事件 [7,10]。在更高层次上，患者旅程可以描述为一系列临床里程碑，例如症状出现、诊断和开始治疗，以及它们之间的转变。患者是否能够达到某个里程碑以及何时达到，取决于多种因素，例如动机、感知障碍和外部触发因素。对里程碑及其相关因素进行建模，可以显示患者旅程在走向成功治疗的过程中会在哪些环节停滞。

改善患者体验需要考察其实际发生的患者旅程。然而，并不存在记录患者旅程的中央系统。因此，数据可能分散在多个系统中（例如不同的医院、诊所和医生报告中）。有些步骤可能完全没有记录（例如与心理学家进行的一次会谈）。这使患者自身成为最了解情况的利益相关者，也因此成为患者旅程证据的重要来源。向患者收集这些证据并非易事，以下示例对此进行了说明。

2.2 示例说明

重症肌无力（MG）是一种自身免疫性疾病，抗体会攻击神经肌肉接头，损害神经向肌肉传递的信号。患者会出现肌无力，活动时很快感到疲惫；症状可能遍及全身（全身型重症肌无力），也可能只影响特定肌群，例如眼周肌肉，或负责吞咽、说话、咀嚼和呼吸的延髓肌。治疗方法包括通过抑制乙酰胆碱酯酶来缓解症状，或抑制免疫攻击。下面，我们将介绍一位虚构患者因重症肌无力而受到影响的患者旅程，作为贯穿全文的示例。

一位患者的旅程。Lena 在汉堡教书。大多数日子里，她都还好，直到下午一点左右；最后两节课时，她的声音变得微弱而带鼻音，学校里还传开了她午餐时喝了酒的传言。

事情始于一种睡眠也无法消除的疲倦，那是在出现任何明确症状的一年多以前。她最先能明确指出的症状，是吃到一半时下颌就感到疲劳。一位耳鼻喉科医生检查后发现她的声带正常，认为这是紧张所致，便将她转介至言语治疗；经过六次治疗后，言语治疗师第一个指出情况有些不对劲。到2023年初，她喝水时会呛咳，体重也减轻了七公斤；一次严重发作让她去了急诊科，但当晚就被送回家。她被转介至神经内科，并等了四个月。最常见抗体的检测结果为阴性，神经刺激检查也正常，但血液检查显示，她于2019年确诊的桥本甲状腺炎治疗不足。甲状腺功能减退症足以解释所有症状，于是医生增加了她的左甲状腺素剂量，甲状腺水平也恢复正常。但其他症状依旧。五个月后，她自费看了一位私人专科医生，对方检测了第二种抗体——肌肉特异性激酶（MuSK）；结果呈阳性，时间是2023年11月，也就是她首次提起这些症状近两年后。“我哭了，然后愤怒了大约六个月。”她保留了出院信，却没有留下此前那些未能诊断出病情的就诊记录。

吡啶斯的明让她出现痉挛，三个月后便停用了。类固醇有效，但她体重增加了11公斤，失去了睡眠，脸型也发生了变化。随后开始使用硫唑嘌呤。2025年2月，一名值班医生用抗生素治疗尿路感染，但两天内她就无法抬起头离开枕头：在医院住了9天，并接受了免疫球蛋白输注。经过两轮利妥昔单抗治疗后，她每周回去工作四天，状态比2022年以来任何时候都好。

3 方法论

3.1 挑战与要求

这个贯穿全文的示例展现了三个关键挑战。我们将患者报告的每个临床事件，例如一次就诊、一个诊断或一项治疗，称为一个 事实。术语 事件 指事件日志（例如 OCEL）中的一个事件。

C1：开放世界。数据中存在不确定性。只有通过所报告的内容才能了解一段病程，而报告的内容并不完整，也可能发生变化。Lena 保留了出院信，但没有保留早期就诊的文档记录；她最初的症状发生在“至少一年前”，而后续报告可能会纠正早期报告。未被报告的事实仍可能发生过，只有明确的陈述才能说明某件事并未发生。

C2：自然语言。病程以患者自己的话语讲述。同一个概念可能用不同的词语、以不同的详细程度进行描述，例如，用“她的声音变细并转为鼻音”描述言语症状，或用“类固醇”指代某种具体药物。必须判断一段陈述所指的是哪个概念。

C3：复杂领域。一段病程涉及多种疾病状况、治疗和利益相关者，并记录于异质的来源中。Lena 的病程贯穿重症肌无力、桥本甲状腺炎和尿路感染，涉及耳鼻喉科医生、言语治疗师、急诊科、神经内科和私人专科医生，并记录在她的叙述和一封出院信中。其中的事实相互影响。患者、临床医生和研究人员从不同的视角审视同一段病程。

根据挑战 C1–C3，我们推导出六项要求。R1、R2 和 R4 针对开放世界（C1），R3 针对自然语言（C2），R5 和 R6 则针对复杂领域（C3）。C3 要求进行多视角分析，例如使用 OCEL。

R1：可追溯证据（C1）。随着新证据不断完善、纠正或反驳早先的报告，人们对一段旅程的了解也会发生变化。因此，每项事实都必须能够追溯到其所依据的证据，以及该事实被报告或修订的时间。

R2：不确定时间（C1）。患者讲述自己的旅程时，并不会按照时间顺序进行。他们会在不同时间之间跳转，给出时间范围（“到 2023年初”）或相对时间（“五个月后”），报告持续时间但不提供日期，或者完全不说明时间。我们只知道 Lena 早期那段经历——从最初的疲倦到接受言语治疗——的叙述顺序。然而，OCEL 要求为每个事件提供时间戳。因此，必须区分叙述顺序、时序和时间戳，而且不能丢失那些无法确定时间的事实。

R3：规范化术语（C2）。患者报告的概念必须规范化为临床术语，同时保留患者自己的词语。

R4：未发生（C1）。未发生的事情可能与发生过的事情同样重要。患者表示自己从未接受过胸腺切除术，所传达的信息与未提及胸腺切除术不同。计划就诊与实际发生的就诊也有所不同。因此，必须呈现否定事实和意向事实，并将其与已发生的事实区分开来。

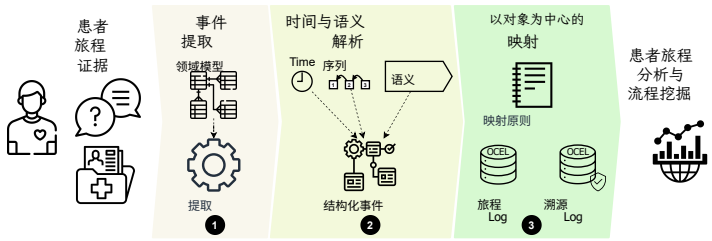


图 1. 从患者旅程构建以对象为中心的事件日志的逐步方法。

R5：多种视角（C3）。需要细粒度的事实来保留个体历史。需要高层抽象来比较不同患者的病程，并识别共同模式、瓶颈和延误。因此，必须能够从多个视角（例如疾病状况和治疗）以及多个抽象层次分析病程。

R6：相关事实和因素（C3）。患者会将事实彼此关联。耳鼻喉科诊断促成了言语治疗，而危机发生在使用抗生素之后。患者在旅程中的进展也可能取决于一些因素，例如障碍因素、激励因素和诱因（如专科医生把关或对副作用的恐惧）。OCEL 2.0 仅通过事件共同涉及的对象，隐式地关联事件 [3]。因此，必须表示关系和因素，并区分患者报告的关联与已确立的临床因果关系。

3.2 从患者旅程创建 OCEL 的方法

图 1 展示了我们方法的各个步骤。我们假设，患者旅程的证据来自访谈、问卷和临床记录，也可能由人工智能代理收集。步骤 1 从这些证据中提取事实。领域模型规定要提取哪些事件类型、关系和临床概念。每项事实都会注明其提取自哪项证据，并保留其修订历史（R1）；同时标记为已发生、已否定或计划发生（R4）。步骤 2 对提取出的事实进行时间和语义解析：将报告的时间转换为日期范围和时间关系（R2），并将患者的措辞与临床术语相匹配（R3）。步骤 3 按照一组映射原则，将结构化临床事件（即事实）映射到两个 OCEL：一个患者旅程日志和一个溯源日志。在患者旅程日志的基础上，里程碑和转变提供了高层抽象。随后，可以使用对象中心流程挖掘技术分析生成的日志。

3.3 事件提取与领域建模

为了从患者旅程证据中提取结构化信息，我们需要明确提取什么。这通过领域模型来实现。领域模型定义了需要从证据中提取的事件类型、事件间关系、标准术语以及

其他相关实体。图 2 展示了一个基于我们行业合作伙伴所使用模型的 UML领域模型。

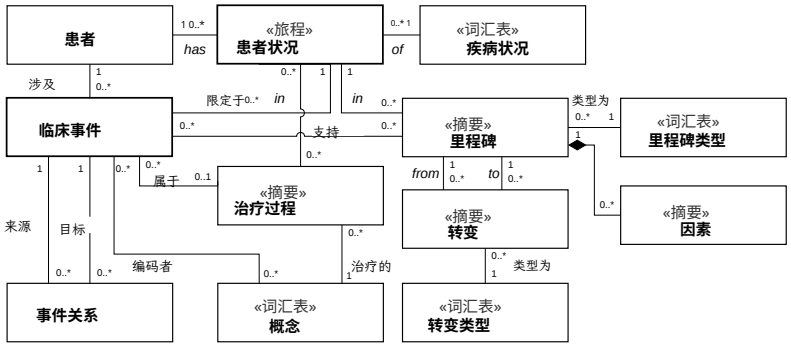


图 2. 患者旅程的领域模型。«词汇表»：共享代码；«摘要»：根据事实得出。

我们通过贯穿全文的示例来说明该模型。Lena 是一名患者，报告了几种疾病状况，即重症肌无力、桥本甲状腺炎和尿路感染。患者与疾病状况的每种组合称为患者状况，并被视为一段病程。Lena 报告的所有内容，例如耳鼻喉科就诊或言语治疗，都是临床事件，我们称之为事实。临床事件属于 Lena，也属于她任意数量的病程（例如，一次就诊中讨论了几种疾病状况），并由共享词汇表中的概念编码。临床事件之间可以通过事件关系相互关联。每种关系都有一个来源、一个目标和一种类型，类型可以是因果关系或时间关系。例如，抗生素会导致重症肌无力危象，而言语治疗发生在耳鼻喉科就诊之后。根据临床事件，可以得出每段病程的摘要。Lena 的吡啶斯的明治疗构成一个治疗过程。MuSK检测结果呈阳性，支持了达到诊断这一里程碑，其里程碑类型为诊断。此前还有一个误诊里程碑，因此两者通过转变相连。转变表示时序，而非因果关系。因素记录了促成或阻碍达成某一里程碑的原因。例如，医生不熟悉重症肌无力以及疾病罕见性，都阻碍了及时诊断和治疗。

3.4 时间与语义分辨率

叙述会在时间中来回跳转，患者会用自己的词语描述临床概念。此步骤的目标是根据时间、序列和语义解析事实。

时间。 报告顺序并不等于发生顺序。因此，除非报告了确切日期，否则事实发生的时间会以带有精度的日期范围表示。例如，“到2023年初”会转换为从2023年1月到3月的范围，精度为月精度。保留报告的持续时间，因此

如果开始时间已知，就可以推导出事实的结束时间戳（例如，吡啶斯的明“在三个月后停用”（R2））。
序列。如果日期范围不足以确定事实的先后顺序，大语言模型（LLM）便会从叙述序列中提取时间关系。这对于判断哪些事实同时发生尤其有用（例如，确定联合治疗方案（R2））。在这个示例中，Lena 报告称，硫唑嘌呤是在类固醇“之后”使用的。只有时间关系才能说明硫唑嘌呤是取代了类固醇，还是在其基础上新增的治疗。

语义。症状名称和治疗名称等临床术语通过与词汇数据库进行模糊匹配来确定，同时保留患者自己的用词（R3）。例如，“她的声音变弱并带有鼻音”会匹配到一种言语症状。为此可以使用任何相关的共享词汇表。在本研究中，我们使用行业合作伙伴正在使用的自定义词汇表。

3.5 以对象为中心的映射

此步骤将结构化事实映射到 OCEL 2.0。
M1：事实是对象，其发生情况是事件（R2、R4）。每个事实都会成为一个对象，其类型为领域模型中的事件类型，例如治疗或医疗专业人员就诊。只有当事实确实发生且时间已知时，才会创建 OCEL 事件：一个时间点事实对应一个事件；一个时间段事实对应一个开始事件，如果已知结束时间或可根据报告的持续时间推知结束时间，则还会对应一个结束事件。时间戳取报告日期范围的起始时间；该范围及其精度会作为事件属性保留。在这个示例中，Lena 的吡啶斯的明治疗会生成一个发生在2023年11月的开始治疗事件，以及一个三个月后发生的结束治疗事件。她就诊耳鼻喉科医生的时间未知，而计划中的门诊复诊尚未进行。因此，这些内容会作为没有事件的对象保留在日志中。报告变化的事实，例如停止治疗或症状改善，会成为其所改变对象上的事件。停止抗生素会成为治疗过程上的一个停止治疗事件，而疾病状态发生变化则会成为旅程的疾病状态属性的新值。里程碑的映射方式与事实相同。每个里程碑都会成为一个对象；已达到且时间已知的里程碑也会生成一个事件，例如达到诊断。未达到的里程碑（明确表示未达到）会保留为没有事件的对象，而未知的里程碑（未曾提及）则不会出现在日志中。里程碑的因素会成为其属性。

M2：事实之间的关系转化为对象之间的关系（R6）。由于 OCEL 2.0 仅通过对象关联事件，因此两个事实之间的关系会转化为其对象之间的 O2O 关系，并将关系类型作为限定符。例如，抗生素治疗会导致重症肌无力危象。限定符用于命名关系的类型。由于 O2O 关系不携带属性，因此不会保留该关系是由患者报告的，还是通过临床方式确定的这一信息。若要保留这一信息，就需要为每个关系创建一个对象。里程碑之间的转变也会以相同的方式转化为 O2O 关系。

M3：视角成为对象类型（R5）。视角是一种实体，我们希望像传统流程挖掘中的案例概念一样，追踪它随时间的行为

。每个视角都成为一种对象类型，其包含的每个实体都成为一个对象。事件与事实相关联，因为事实记录了该事件；而通过该事实，事件又与其所属每个视角中的对象相关联。因此，可以从不同视角分析同一批事件。在这个贯穿全文的示例中，患者、病程和治疗过程都是视角：Lena 开始使用硫唑嘌呤这一事件同时属于她本人、她的重症肌无力旅程以及她的硫唑嘌呤疗程。**M4：术语经过标准化，措辞予以保留（R3）**。每个事实都将标准化术语和患者自己的措辞保留为属性。诸如“其他治疗”这样的兜底代码不会用作概念，因为它们会将彼此无关的事实合并。相反，概念由标准化措辞推导得出，或者省略。
M5：记录已发生的事件，并将报告事件保存在单独的日志中（R1）。每个 OCEL 属性值都只有一个时间戳。因此，我们创建了两个日志。历程日志 使用事实发生的时间。溯源日志 使用事实被报告的时间。它包含每一项事实，无论该事实是否在患者旅程日志中对应一个事件，并记录事实所依据的消息、问卷答案和上传文档，以及每项事实被提出或修订的时间。在这两种日志中，每项事实都指向提取该事实所依据的证据。

3.6 患者旅程分析

可以从每个视角分析 OCEL（M3）。在我们的场景中，这些视角分别是患者、旅程和治疗过程。以案例为中心的日志只提供一个视角，例如患者视角。提高抽象层次也无法消除这一局限。即使某项治疗事件的治疗过程与另一项并行，它仍然是患者的事件；对重症肌无力的诊断和对哮喘的诊断也仍属于同一种活动。里程碑和转变提供了一个抽象层次，可据此比较不同患者的病程。对里程碑进行流程挖掘，可以展示最常见的路径，以及其中的瓶颈、延误和流失。

4 评估

我们将该方法应用于一组合成患者的病程。这些数据来自 mama health 合成生成的证据，模拟了真实世界中的患者旅程；由于需要保护个人可识别健康数据的隐私，这些真实患者旅程无法公开。数据包含 280 名患有重症肌无力（MG）的合成患者（另有 4 名患者因缺少病程摘要而未纳入）。他们报告了 13,684 条事实和 2,267 个里程碑。基于这些数据，我们创建了两份日志。患者旅程日志将每条事实作为一个对象，但有 72 条疾病状态变化例外，这些变化在其旅程中作为事件（M1）。在 13,684 条事实中，有 7,346 条发生于已知时间，由此产生 8,364 个事件，例如治疗的开始和结束。其余 6,338 条事实（46.3%）是被否定的、计划中的，或缺少可用时间信息，因此仍作为对象，不构成事件。加上里程碑（2,000 个事件）、280 名患者、800 段旅程和 747 个治疗过程，患者旅程日志共包含 17,706 个对象和 10,364 个事件，涉及 30 种活动。溯源日志再次收录了全部 13,684 条事实，每条事实均在其报告时间被断言或修订，并收录了 12,254 条消息，

表1. 要求与原则；● ：通过构造满足，◐ ：部分满足

要求	在 OCEL 中	Met	证据
R1 可追溯证据	证据引用，溯源日志	● M5	全部[13],684 个事实引用了其证据。
R2 不确定 time	日期范围作为属性、无日期事实作为对象	◐ M1	956 个事实（7.0%）没有已确定时间仍然是对象。
R3 规范化术语体系	术语和词语作为属性	● M4	所有术语和措辞均予以保留。
R4 非发生	没有事件	● M1	5,377项事实（39.3%）已被否定或属于预期。
R5 多种视角	视角作为对象类型	● M3	来源链接到某一疾病状况与其旅程相关。
R6 相关事实和因素	O2O 关系，作为属性的因素	◐ M2	保留了 7,184 个关系，但未保留这些关系是否由患者报告或属于临床信息。

2,795 个问卷答案，以及这些消息和答案所依据的 446 份文档。该日志包含 29,179 个事件和 30,553 个对象。复现包可在 GitHub [3] 上找到。

要求。 如果映射后，日志（或对于 R1，溯源日志）中仍保留要求所描述的信息，且这些信息可供分析，则该日志满足此项要求。表1说明了日志如何保留每项要求涉及的信息，以及该项要求是否得到满足。如果某项映射原则保证每个事实都符合某项要求，则该要求按构造满足，因此我们的检查只确认实现是否应用了该原则。我们逐项检查日志中的要求，并将计数与映射引擎报告的源数据计数进行比较。

两项要求仅部分满足。R2 受到 OCEL 本身的限制。一个事件只有一个时间戳，因此，如果某个日期仅精确到月份，就会被设为该月的第一天。因此，患者的直接跟随对中有 31.0% 共享同一时间戳，而 PM4Py 会以任意顺序排列它们。R6 未保留关系依据（M2）。其余四项要求按构造满足。

患者旅程。 我们使用 PM4Py 将患者的历程日志扁平化。结果是一个以案例为中心的日志，包含 280 个案例和全部 10,364 个事件，且每个事件恰好出现一次。仅保留里程碑后，该日志与行业合作伙伴此前分析的历程摘要日志相符。其中 15 种里程碑类型的 2,000 个事件构成了 204 个变体。PM4Py 发现的直接跟随图见图 3。常见的历程指标也由同一日志得出。从症状出现到诊断的中位时间为 12.7 个月（四分位距 7.0–18.3，$n = 251$）。误诊患者的这一时间为 19.6 个月（$n = 57$），其他患者则为 10.6 个月（$n = 194$）。

[3] https://github.com/mama-neven/OCEL-patient-journeys-replication-package

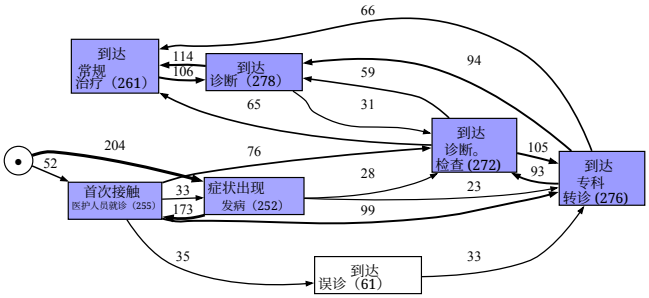


图 3。 以患者为案例的常规治疗前里程碑直接跟随图（PM4Py）。隐藏少于 23 个配对的边。

我们的行业合作伙伴确认，这些结果与其自行对历程摘要所做的分析相符。

对象中心流程挖掘。 患者旅程和治疗过程各自都是独立的对象。在历程日志中，280 名患者中有 211 名报告了除 MG 之外的合并症旅程。以患者为案例时，MG 诊断和哮喘诊断属于同一个活动，即 ReceiveDiagnosis。以旅程为案例时，与上层患者的关联会丢失，因此，38.5% 的未与旅程关联的事实事件也会丢失。对象中心日志则同时保留了两者。我们使用 PM4Py，根据原样的历程日志发现了图 4 中的对象中心直接跟随图。旅程和治疗过程共享治疗活动，但路径并不相同。477 个旅程以诊断开始，而一个治疗过程以治疗开始。治疗过程结束后即停止：其路径从不从 EndTreatment 返回到 StartTreatment（0 个配对）。旅程的路径返回了 80 次，涉及 67 个旅程。每次开始的治疗都属于同一旅程中的另一个治疗过程（80 次中全部如此）。

因此，图中展示了治疗切换，尽管没有任何活动以此命名。以患者为案例时，直接跟随图中同一条边出现了 114 次。它无法区分切换、重新开始，或针对另一种疾病状况的治疗。

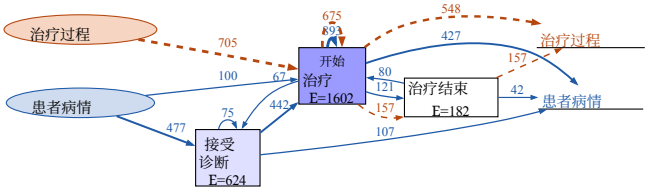


图 4。 历程日志（PM4Py）的对象中心直接跟随图（OC-DFG），仅限于诊断、治疗开始和治疗结束。实线：病程；虚线：治疗过程。

讨论。 仍有一些信息丢失。具体而言，53 个已发生的事实记录了持续时间，却没有结束事件。PM4Py 丢弃了 6,771 个没有事件的对象，以及以其中一个对象为终点的 4,712 个 O2O 关系。要分析哪些事情没有发生，必须直接查阅日志。医疗服务提供者作为就诊的属性予以保留。若要采用医疗服务提供者视角，则需要提供者身份信息，而数据中并不包含这些信息。有效性威胁。这些数据是合成的，可能与领域模型共享某些假设。该队列并未严格留出。所有患者均因重症肌无力而入选，且里程碑和治疗过程仅针对重症肌无力定义。我们未根据独立的真实基准评估其忠实度。

5 相关工作

Munoz-Gama 等人 [9] 描述了医疗保健流程的特征，并确定了分析这些流程所面临的挑战。Arias 等人 [1]和 Shafei 等人 [13]将患者旅程分析视为客户旅程映射问题，并使用流程挖掘分析患者旅程。两者都从信息系统中构建事件日志，并采用单一案例概念。相比之下，我们从患者报告的证据出发，创建以对象为中心的事件日志。

以对象为中心的流程挖掘已应用于 MIMIC-IV，该数据库遵循 OMOP 通用数据模型 [11], ，包括一致性检查 [15]。Bala 等人 [2] 使用大语言模型生成事件日志，并表明不正确的医学术语会限制结果。这些工作从结构化临床数据库出发。在这些数据库中，事实（即领域中的事件）既不会缺少时间，也不会未能发生，并且之后不会被修订。

关于事件日志的创建，Seidel 等人 [12] 概述了相关挑战，Fahland 等人 [5] 提出了一个核心模型，并讨论了其设计空间。其他方法通过目标驱动的方法 [17], ，或扩展 PM^2 方法论 [8]，从支持 SAP ERP [4],的数据库中提取日志。Jans 和 Soffer [6] 指出，创建事件日志涉及许多决策；Swevels 等人 [14] 则提出了流程挖掘感知型领域模型。所有这些方法都以信息系统记录的数据为起点。我们的映射原则以患者报告的证据为起点，并保留没有时间戳或并未发生的事实。

6 结论

本文探讨了如何从患者旅程中创建以对象为中心的事件日志。我们从三个挑战中推导出六项要求，并提出了一种分步方法。其映射原则是将每项事实都保留为一个对象，并且只有在事实发生时间已知时才创建事件。我们将该方法应用于一个由280名重症肌无力患者构成的合成患者队列。生成的日志符合 OCEL 2.0，并保留了46.3%没有对应事件的事实。它们还将以案例为中心的日志会合并的治疗过程区分开来。R2和R6仅部分满足。R2受到单一时间戳的限制，R6则受到简单对象间关系的限制。未来工作将把该方法应用于真实患者旅程和其他疾病状况，并依据真实基准评估提取和解析步骤。

参考文献

1. Arias, M., Rojas, E., Aguirre, S., Cornejo, F., Munoz-Gama, J., Sepúlveda, M., Capurro, D.：通过流程挖掘映射医疗保健中的患者旅程。《国际环境研究与公共卫生杂志》**17**(18), 6586 (2020) 2. Bala, S., Soleimani, Z., Weske, M.：用于医疗保健流程挖掘的人工智能辅助事件日志生成。载于：BPM。LNCS，第16936卷，第259–276页。Springer (2026) 3. Berti, A., Koren, I., Adams, J.N., Park, G., Knopp, B., Graves, N., Rafiei, M., Liß, L., Unterberg, L.T.G., Zhang, Y., 等：OCEL（对象中心事件日志）2.0规范。arXiv预印本 arXiv:2403.01975 (2024) 4. Berti, A., Park, G., Rafiei, M., van der Aalst, W.M.P.：从支持 SAP ERP 的数据库中提取对象中心事件数据的通用方法。《智能信息系统杂志》**61**(3), 835–857 (2023) 5. Fahland, D., Montali, M., Lebherz, J., van der Aalst, W.M.P., 等：迈向对象中心事件数据（OCED）的简单且可扩展标准——核心模型、设计空间与经验教训。CoRR **abs/2410.14495** (2024) 6. Jans, M., Soffer, P.：从关系数据库到事件日志：对质量的影响。载于：业务流程管理研讨会。LNBIP，第308卷，第588–599页。Springer (2017) 7. McCarthy, S., O'Raghallaigh, P., Woodworth, S., Lim, Y.L., Kenny, L.C., Adam, F.：用于将质量融入医疗服务改革的整合式患者旅程映射工具。《决策系统杂志》**25**(sup1), 354–368 (2016) 8. Miri, N., Khayatbashi, S., Zdravkovic, J., Jalali, A.：OCPM^2：扩展流程挖掘方法论以提取对象中心事件数据。载于：BP-MDS/EMMSAD@CAiSE。LNBIP，第558卷，第123–140页。Springer (2025) 9. Munoz-Gama, J., 等：医疗保健流程挖掘：特征与挑战。《生物医学信息学杂志》**127**, 103994 (2022) 10. Oben, P.：理解患者体验：概念框架。《患者体验杂志》**7**(6), 906–910 (2020) 11. Park, G., Lee, Y., Cho, M.：通过对象中心流程挖掘增强医疗保健流程分析：将 OMOP 通用数据模型转换为对象中心事件日志。《生物医学信息学杂志》**156**, 104682 (2024) 12. Seidel, A., Weske, M., Montali, M., Rivkin, A., Reichert, M., van der Werf, J.M.E.M., 等：对象中心流程管理：研究宣言。《信息系统》**141**, 102728 (2026) 13. Shafei, I., Karnon, J., Crotty, M.：医疗保健中的流程挖掘与客户旅程映射：提升卒中康复中的以患者为中心的护理。《数字健康》**10** (2024) 14. Swevels, A., Zerbato, F., Fahland, D.：情境中的流程挖掘：扩展领域数据模型以进行迭代分析。LNCS，第16558卷，第195–212页。Springer (2026) 15. Tripathi, A., Aneesh, Shivam, Y., Pandey, S., Vyas, A., Vyas, O.P.：利用 MIMIC IV 探索对象中心流程挖掘：发掘医疗保健领域的洞见。载于：CAiSE Workshops，LNBIP，第521卷，第360–372页。Springer (2024) 16. Vesinurm, M., Sylgren, I., Bengts, A., Torkki, P., Lillrank, P.：患者旅程中断的概念分析：整合照护的障碍。《整合照护杂志》**32**(5), 37–58 (2024) 17. Wei, J., Ouyang, C., Wang, Y., Huang, L.：Dirigo：一种为对象中心流程提取事件日志的方法。《数据与知识工程》**160**, 102485 (2025)