Journal of Health Informatics

J Health Inform. 2026;18:e1795
DOI: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1795

ARTIGO COMPLETO
CBIS 2026

Mineração de processos aplicada à jornada assistencial do câncer colorretal: análise comparativa de desfechos clínicos e impacto econômico em operadora de saúde suplementar

Process mining applied to the colorectal cancer care journey: comparative analysis of clinical outcomes and economic impact in a private health insurance provider

Gabriela Sanches Bertoli[1] ✉, Marcelo Rosano Dallagassa[1], Andrea Wendt[1], Eduardo Rocker Ramos[1], Sergio Ossamu Ioshii[1]

ORCID: 0009-0004-8886-7772 · 0000-0003-0570-4405 · 0000-0002-4640-2254 · 0009-0004-6998-1027 · 0000-0002-7871-4463

[1] Programa de Pós-Graduação em Tecnologia em Saúde (PPGTS), Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba (PR), Brasil.

FICHA DO ARTIGO

TIPO
Artigo completo · CBIS 2026

IDENTIFICADOR
e1795

DOI
10.59681/2175-4411.v18.iEspecial.2026.1795

RECEBIDO
17/05/2026

APROVADO
20/07/2026

PROCESSO DE AVALIAÇÃO
Double Blind Review

EDITOR RESPONSÁVEL
Juliano de Souza Gaspar

ACESSO
Aberto, CC BY 4.0

AUTOR CORRESPONDENTE
Gabriela Sanches Bertoli
gabibertoli@yahoo.com.br

LICENÇA

Este é um artigo publicado em acesso aberto sob a licença Creative Commons CC BY 4.0, que permite uso, distribuição e reprodução em qualquer meio, desde que o trabalho original seja corretamente citado.

CC BY 4.0

RESUMO

Objetivo: Comparar, por mineração de processos, a jornada assistencial, os desfechos clínicos e o custo entre pacientes com câncer colorretal de início jovem (EOCRC, até 50 anos) e convencional (LOCRC, acima de 50 anos) em operadora de saúde suplementar. **Método:** Coorte histórica retrospectiva de 178 pacientes (2014 a 2024; CID-10 C18 a C20), relatada segundo o STROBE; a jornada foi minerada sobre log deduplicado; a sobrevida por Kaplan-Meier e Cox, o custo por Mann-Whitney (deflação IPCA) e a conformidade frente à SBOC. **Resultados:** A jornada mostrou-se fragmentada (96,0% de trajetórias exclusivas), com quimioterapia em 89,9% do custo. A sobrevida em 60 meses foi maior no EOCRC (87,8% contra 62,7%; p=0,4326), sob poder de 15,4%; a idade associou-se ao óbito (p=0,0212). O custo mediano foi 1,52 vez maior no LOCRC (p=0,1662). **Conclusão:** A mineração evidenciou jornada fragmentada; as diferenças favoráveis ao EOCRC não foram confirmadas estatisticamente, sem constituir ausência de efeito.

Descritores: Neoplasias Colorretais; Mineração de Dados; Saúde Suplementar.

ABSTRACT

Objective: To compare, through process mining, the care pathway, clinical outcomes, and economic impact of early-onset (EOCRC, up to 50 years) versus late-onset (LOCRC, over 50 years) colorectal cancer patients in a supplementary health operator. **Method:** Retrospective historical cohort of 178 patients (2014 to 2024; ICD-10 C18 to C20), reported per STROBE; the pathway was mined over a deduplicated event log; survival by Kaplan-Meier and Cox, cost by Mann-Whitney (IPCA deflation), and conformance against SBOC 2026. **Results:** The pathway was fragmented (96.0% unique trajectories), with chemotherapy concentrating 89.9% of costs. Sixty-month survival was higher in EOCRC (87.8% vs. 62.7%; p=0.4326), under 15.4% power; continuous age was associated with death (p=0.0212). Median real cost was 1.52 times higher in LOCRC (p=0.1662). **Conclusion:** Process mining revealed a fragmented pathway; differences favoring EOCRC were not statistically confirmed, without constituting absence of effect.

Keywords: Colorectal Neoplasms; Data Mining; Supplemental Health.

CBIS 2026 XXI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE

Artigo apresentado no XXI Congresso Brasileiro de Informática em Saúde (CBIS 2026), realizado de 23 a 25 de setembro de 2026, em Brasília (DF), e publicado na seção CBIS 2026 do Journal of Health Informatics.

COMO CITAR ESTE ARTIGO: Bertoli GS, Dallagassa MR, Wendt A, Ramos ER, Ioshii SO. Mineração de processos aplicada à jornada assistencial do câncer colorretal: análise comparativa de desfechos clínicos e impacto econômico em operadora de saúde suplementar. J Health Inform 2026;18:e1795. doi: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1795

1 INTRODUÇÃO

O câncer colorretal (CCR) é a terceira neoplasia mais incidente e a segunda causa de morte por câncer no mundo, com cerca de 1,9 milhão de casos novos e 930 mil óbitos estimados em 2022 [1]. No Brasil, o Instituto Nacional de Câncer (INCA) projetou aproximadamente 45.630 casos novos para 2023 e 2024, com tendência de crescimento para cerca de 53.800 casos anuais no triênio 2026 a 2028 [2]. Na saúde suplementar, o CCR está entre as patologias de maior custo assistencial, dada a intensidade dos regimes terapêuticos [3].

Um fenômeno epidemiológico de incidência crescente é o câncer colorretal de início jovem (*early-onset colorectal cancer*, EOCRC), diagnosticado antes dos 50 anos, em contraposição ao padrão convencional de diagnóstico após essa idade (*late-onset colorectal cancer*, LOCRC). A incidência de EOCRC cresce a taxas de 1,5% a 3,6% ao ano em múltiplos países, incluindo o Brasil [4,5]. Fatores dietéticos e de estilo de vida têm sido implicados na sua etiologia [6], enquanto a incidência nas faixas etárias mais velhas se estabiliza ou declina em razão dos programas de rastreamento [7,8].

A análise da jornada assistencial do CCR em bases administrativas de operadoras enfrenta desafios metodológicos: ausência de estadiamento estruturado pela classificação Tumor-Nódulo-Metástase (TNM), subregistro de óbito, heterogeneidade temporal dos custos e fragmentação dos registros entre múltiplos prestadores. A mineração de processos (MP), que reconstrói trajetórias reais a partir de logs de eventos dos sistemas de informação [9], permite caracterizar a jornada, identificar gargalos e integrar as dimensões clínica e econômica do cuidado [10,11]; um mapeamento sistemático de 270 estudos evidenciou as oportunidades e os desafios dessa aplicação em saúde, incluindo a qualidade dos dados, a privacidade e a construção do log de eventos [12]. No Brasil, sua aplicação em oncologia é incipiente, sem estudos que integrem jornada, sobrevida e custo em CCR na saúde suplementar sob recorte etário EOCRC/LOCRC. O objetivo deste trabalho é comparar, por MP, a jornada assistencial, os desfechos clínicos e o impacto econômico entre EOCRC e LOCRC em uma operadora de saúde suplementar.

2 MÉTODOS

Estudo observacional, longitudinal e retrospectivo, de abordagem quantitativa, relatado segundo as diretrizes STROBE (*Strengthening the Reporting of Observational Studies in Epidemiology*) [13]. Utilizaram-se dados administrativos anonimizados da Unimed Paraná (registro na Agência Nacional de Saúde Suplementar, ANS, nº 312720), de janeiro de 2014 a dezembro de 2024, referentes a pacientes com CCR confirmado (Classificação Internacional de Doenças, 10ª revisão, CID-10: C18, C19 e C20). O estudo foi aprovado pelo Comitê de Ética em Pesquisa (CEP) da PUCPR (Parecer nº 7.857.929), em conformidade com as Resoluções do Conselho Nacional de Saúde (CNS) nº 466/2012 e nº 510/2016 e com a Lei Geral de Proteção de Dados (LGPD, Lei nº 13.709/2018). Prevê-se, como etapa subsequente, validação por painel de especialistas oncológicos nos casos de Desvio Severo e de conformidade Indeterminada, conferindo validade clínica aos achados automatizados.

2.1 Definição de grupos e limitações do critério de estadiamento

Os pacientes foram estratificados pela idade no marco diagnóstico (T0): Grupo A, EOCRC (T0 até 50 anos; n=31), e Grupo B, LOCRC (T0 acima de 50 anos; n=147). O T0 foi definido como a data do primeiro procedimento diagnóstico documentado, segundo hierarquia determinística de prioridade clínica: colonoscopia ou retossigmoidoscopia, cirurgia de ressecção, primeiro CID de CCR primário, biópsia e, excepcionalmente, tomografia por emissão de pósitrons associada à tomografia computadorizada (PET-CT). Na ausência de campos TNM estruturados, limitação inerente a bases administrativas, o estadiamento foi inferido por *proxy* terapêutico; o estádio IV foi determinado por critério combinado (CID-10 de sítio metastático C77 a C79 ou uso de agente de doença avançada). Trata-se de aproximação probabilística, sujeita a viés de indicação e a erro de classificação individual. Essa é a principal limitação do critério de agrupamento, assumida de forma explícita e pendente de validação pelo painel de especialistas.

2.2 Construção do log de eventos e parâmetros do algoritmo

O log foi estruturado no padrão XES (*eXtensible Event Stream*), com os elementos *Case ID*, *Activity*, *Timestamp* e atributos de caso. Construiu-se um vocabulário controlado de 23 macroatividades a partir de mais de 8.000 descrições de serviço, priorizando a classe farmacêutica na detecção de quimioterapia; retiveram-se apenas eventos posteriores ao T0 relacionados ao CCR; e aplicou-se deduplicação temporal (mesmo caso, mesma macroatividade e mesmo dia), que corrigiu o fracionamento de faturamento e reduziu o log de 16.485 para 6.168 eventos (queda de 62,6%), preservando custo total e estrutura de variantes. Sobre o log deduplicado, computaram-se, conforme os algoritmos clássicos de MP [14], a frequência de casos e por atividade, as repetições (*loops*), a causalidade por grafo de fluxo direto (*directly-follows graph*, DFG) e a duração média, sob múltiplas perspectivas (macroatividade, trajetória e conformidade segundo a Sociedade Brasileira de Oncologia Clínica (SBOC) [15], desfecho, estádio e modalidade de atendimento) e em dois níveis de abstração (80% e 40%). A MP foi executada na plataforma UpFlux com o algoritmo Fuzzy Miner, o mais empregado em MP na saúde por acomodar a

variabilidade das trajetórias e mitigar o efeito espaguete [12], gerando os mapas nativos e o *conformance checking* por *replay* de *tokens*, com as métricas confirmadas e validadas em Python; a empresa não participou da concepção, análise ou redação do trabalho.

2.3 Análise estatística e econômica

A sobrevida global foi caracterizada primariamente de forma descritiva (distribuição de desfechos, sobrevida em 60 meses com intervalo de confiança de 95% (IC95%) por Greenwood e medianas) e complementada por teste log-rank e regressão de Cox. As curvas de sobrevida seguiram o método de Kaplan-Meier [16]; a associação com a idade contínua foi estimada pelo modelo de Cox [17]; e o pressuposto de riscos proporcionais foi avaliado pelos resíduos de Schoenfeld [18]. Consideraram-se dois cenários: A (óbito formalmente codificado) e D (óbito confirmado, provável ou possível). Diante da ausência de significância estatística, calculou-se formalmente o poder *post-hoc*. Os custos foram deflacionados pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA) [19] evento a evento (base dez/2024) e comparados por Mann-Whitney. A conformidade foi avaliada por comparação das trajetórias (T1 a T5) com as diretrizes da SBOC 2026 [15]. Adotou-se significância de 5%; as análises foram conduzidas em Python (*lifelines* [20], *scipy*).

3 RESULTADOS

3.1 Caracterização da coorte

A coorte reuniu 178 pacientes: EOCRC (n=31; 17,4%) e LOCRC (n=147; 82,6%). No EOCRC houve predominância feminina (58,1%); no LOCRC, masculina (56,5%). A idade mediana ao T0 foi de 44,0 anos no EOCRC (intervalo interquartil, IQR, 38,3 a 46,7) e de 68,2 anos no LOCRC (IQR 59,6 a 73,9). O T0 foi ancorado por colonoscopia ou retossigmoidoscopia em 84,3% dos casos (Tabela 1).

O estadiamento inferido mostrou predomínio de doença avançada, com estádio IV em 38,7% do EOCRC e 36,7% do LOCRC (Tabela 2).

3.2 Mineração de processos: análise multiperspectiva da jornada

O log deduplicado reuniu 6.168 eventos em 177 casos e 17 macroatividades (média de 34,8 eventos por caso). As mais frequentes foram quimioterapia (2.417 eventos; 115 casos), marcador tumoral (1.547; 161), imagem de estadiamento (1.147; 171), colonoscopia (402) e cirurgia de CCR (358). O DFG identificou 100 transições, com a sequência de marcador tumoral para quimioterapia como a mais frequente (801 ocorrências), o que configura um núcleo clínico de vigilância e tratamento. Das 173 variantes distintas entre os 177 casos, 170 (96,0%) são exclusivas de um único paciente, proporção que se manteve antes e após a deduplicação. A quimioterapia concentrou 89,9% do custo do log.

A jornada foi analisada em dois níveis complementares de granularidade, para os dois grupos etários. Na Visão Macro (linha de tratamento), agrupada pela variável TRAJETORIA_SBOC, sintetizam-se as trajetórias terapêuticas (T1, cirurgia curativa; T2, cirurgia com quimioterapia adjuvante; T3, quimioterapia metastática; T4, quimioterapia sem cirurgia registrada; T5, sem tratamento registrado). Na Visão Micro (detalhamento operacional), agrupada por MACROATIVIDADE, detalha-se o fluxo entre as cinco etapas assistenciais núcleo (colonoscopia, imagem de estadiamento, marcador tumoral, cirurgia e quimioterapia). Dada a granularidade de nível-dia do log, reportam-se os tempos médios de transição entre etapas, em dias.

Na Visão Macro (Figura 1), as trajetórias apresentaram durações médias distintas entre os grupos. No EOCRC predominaram T2 (1.100,0 dias), T1 (953,1 dias), T3 (530,2 dias) e T5 (88,0 dias). No LOCRC, além dessas, emergiu a trajetória T4 (900,3 dias), ausente no EOCRC, com durações de T3 (1.045,7 dias), T1 (657,5 dias), T2 (843,0 dias) e T5 (608,0 dias). A trajetória de intenção curativa isolada (T1) foi mais longa no EOCRC do que no LOCRC (953,1 contra 657,5 dias). As trajetórias T4 (10 casos) e T5 (13 casos) são o alvo das duas frentes de validação por especialistas.

Na Visão Micro (Figura 2), os grafos das cinco etapas núcleo mostraram que o principal componente do *lead time* é o tempo de transição entre etapas, e não a duração dos procedimentos. O intervalo entre colonoscopia e imagem de estadiamento foi maior no EOCRC do que no LOCRC (cerca de 121 contra 77 dias), indicando atraso mais acentuado entre a suspeita e a conclusão do estadiamento nos pacientes jovens. No LOCRC, destacaram-se sinais de fragmentação do acompanhamento: autorrepetição de colonoscopia (cerca de 113 dias) e de marcador tumoral (cerca de 90 dias), além do laço de retorno de cirurgia para colonoscopia (cerca de 58 dias), compatíveis com reinvestigação e retrabalho diagnóstico no grupo sênior.

3.3 Análise de sobrevida

No Cenário A registraram-se 4 óbitos no EOCRC (12,9%) e 32 no LOCRC (21,8%); no Cenário D, 6 (19,4%) e 40 (27,2%). A maior parte da coorte permaneceu censurada no Cenário D (74,2%). A sobrevida em 60 meses foi maior no EOCRC em ambos os cenários: no Cenário A, 87,8% (IC95%: 65,6 a 96,1) contra 62,7% (IC95%: 49,0 a 73,7) no LOCRC (Figura 3); no Cenário D,

Tabela 1. Critério de definição do T0 por grupo etário

Critério T0	EOCRC n (%)	LOCRC n (%)	Total n (%)
Colonoscopia/Retossigmoidoscopia	24 (77,4)	126 (85,7)	150 (84,3)
Cirurgia de ressecção intestinal	2 (6,5)	9 (6,1)	11 (6,2)
1º CID de CCR primário	2 (6,5)	6 (4,1)	8 (4,5)
Biópsia anatomopatológica	2 (6,5)	5 (3,4)	7 (3,9)
PET-CT oncológico	1 (3,2)	1 (0,7)	2 (1,1)
Total	31 (100)	147 (100)	178 (100)

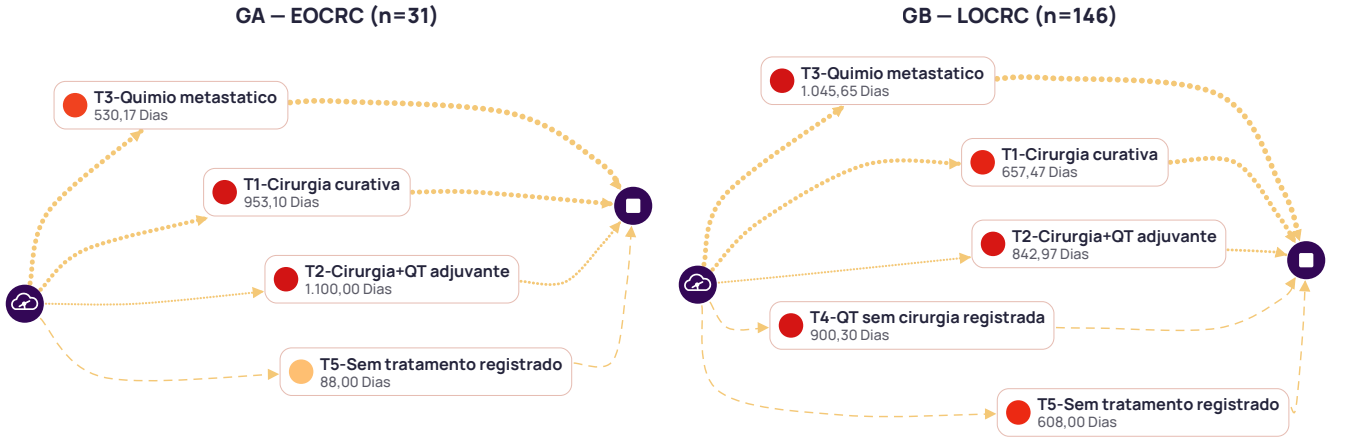
Fonte: Elaborado pelos autores (2026).

Tabela 2. Distribuição por estádio inferido segundo grupo etário

Estádio inferido	EOCRC n (%)	LOCRC n (%)	Total n (%)
I/II	10 (32,3)	45 (30,6)	55 (30,9)
III	6 (19,4)	28 (19,0)	34 (19,1)
IV	12 (38,7)	54 (36,7)	66 (37,1)
II/III ambíguo	0 (0,0)	10 (6,8)	10 (5,6)
Indeterminado	3 (9,7)	10 (6,8)	13 (7,3)
Total	31 (100)	147 (100)	178 (100)

Fonte: Elaborado pelos autores (2026).

Figura 1. Visão Macro da jornada assistencial (variável TRAJETORIA_SBOC): trajetórias terapêuticas na perspectiva de duração média (dias), por grupo etário. GA, Grupo A (EOCRC); GB, Grupo B (LOCRC). O LOCRC inclui a trajetória T4 (quimioterapia sem cirurgia registrada), ausente no EOCRC.



O Grupo B aparece com n=146 (exclui um paciente sem evento pós-T0); os totais de referência do estudo são coorte n=178, EOCRC n=31 e LOCRC n=147.

Fonte: UpFlux Process Mining, 2026.

81,4% contra 57,3%.

A diferença entre as curvas não alcançou significância estatística (log-rank p=0,4326 no Cenário A; p=0,5878 no D), com poder *post-hoc* de 15,4% e 8,9%; a confirmação exigiria um EOCRC de aproximadamente 274 a 652 pacientes, de modo que a ausência de significância não equivale a ausência de diferença real. A regressão de Cox indicou associação da idade contínua com o óbito (razão de risco, *hazard ratio*, HR=1,037 por ano; IC95%: 1,005 a 1,069; p=0,0212 no Cenário A; HR=1,028; p=0,0429 no Cenário D), com riscos proporcionais sustentados (resíduos de Schoenfeld p=0,288 e p=0,433).

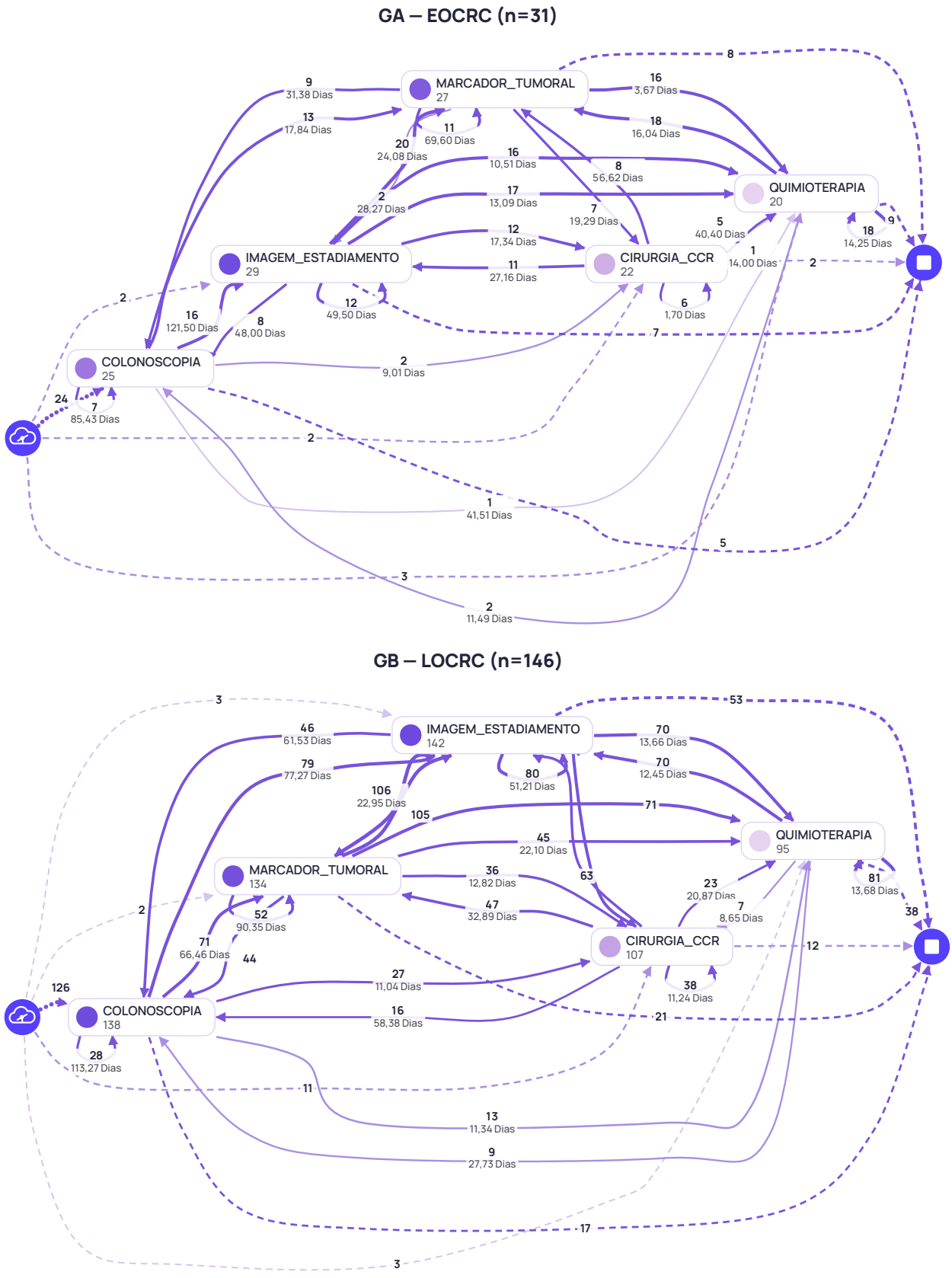
3.4 Análise econômica e conformidade

O custo total real da coorte foi de R$ 50.781.296,52. O custo mediano real por paciente foi de R$ 111.123,64 (EOCRC) e R$ 169.081,94 (LOCRC), razão de 1,52 vez, sem significância estatística (Mann-Whitney U=1.917; p=0,1662), sob alta variabilidade da distribuição (Tabela 3).

O total nominal do LOCRC supera o do EOCRC em 5,54 vezes, diferença explicada majoritariamente pela proporção populacional; por paciente, a diferença é menor. A conformidade plena ou por adjuvância foi de 90,3% (EOCRC) e 85,7% (LOCRC); o Desvio Severo

somou 14 casos (7,9%) e dez casos ambíguos (todos LOCRC) foram classificados como Indeterminado

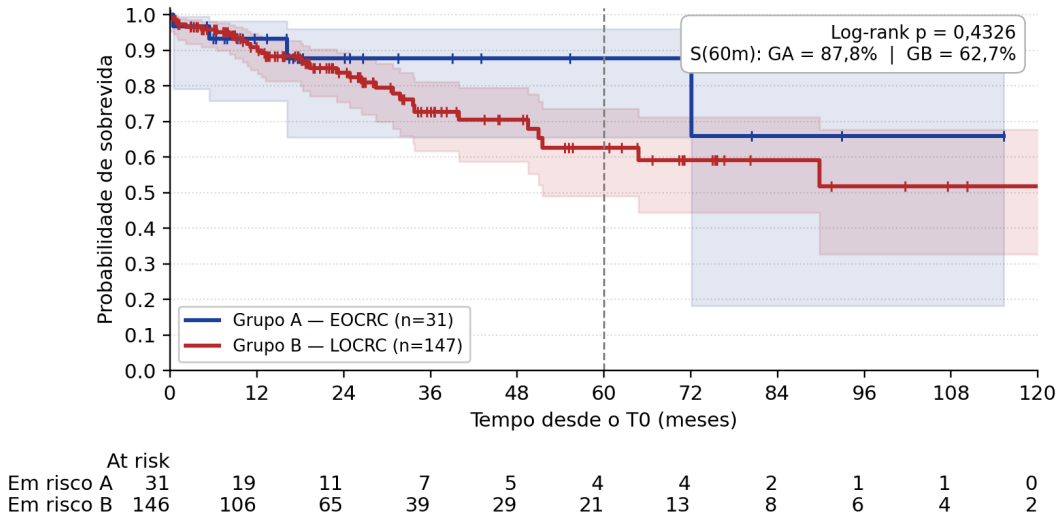
Figura 2. Visão Micro da jornada assistencial (variável MACROATIVIDADE): fluxo entre as cinco etapas núcleo, com frequência de casos e tempo médio de transição em dias, por grupo etário. GA, Grupo A (EOCRC); GB, Grupo B (LOCRC).



Log deduplicado; nas arestas, frequência de casos (número superior) e tempo médio de transição em dias (inferior).

Fonte: UpFlux Process Mining, 2026.

Figura 3. Curva de Kaplan-Meier (Cenário A) por grupo etário, com IC95% e número de pacientes em risco.



Estimador de Kaplan-Meier (lifelines); IC95% por Greenwood; comparação por teste log-rank. Evento: óbito formalmente codificado (Cenário A); censura no último atendimento registrado. Tempo medido em meses desde o T0 (n=178).

Fonte: Elaborado pelos autores (2026).

Tabela 3. Análise econômica por grupo etário (valores reais, deflação pelo IPCA, base dez/2024)

Métrica	EOCRC (R$)	LOCRC (R$)	Razão
Custo total real	7.768.890,96	43.012.405,56	5,54 vezes
Custo mediano por paciente	111.123,64	169.081,94	1,52 vez
Custo médio por paciente	250.609,39	292.601,40	1,17 vez

Total geral: R$ 50.781.296,52. Mann-Whitney U=1.917; p=0,1662.

Fonte: Elaborado pelos autores (2026).

Tabela 4. Conformidade com as diretrizes SBOC 2026 por grupo etário

Categoria	EOCRC n (%)	LOCRC n (%)	Total n (%)
Conforme Pleno	22 (71,0)	98 (66,7)	120 (67,4)
Conforme Adjuvância	6 (19,4)	28 (19,0)	34 (19,1)
Desvio Severo	3 (9,7)	11 (7,5)	14 (7,9)
Indeterminado (pendente)	0 (0,0)	10 (6,8)	10 (5,6)
Total	31 (100)	147 (100)	178 (100)

Fonte: Elaborado pelos autores (2026).

(Tabela 4).

4 DISCUSSÃO

A coorte reflete o padrão do CCR na saúde suplementar: predominância do LOCRC (82,6%) e presença de EOCRC em 17,4%, acima da estimativa nacional do INCA de 10% a 12%, possivelmente pela estrutura etária da carteira de beneficiários e pelo acesso facilitado à investigação diagnóstica.

A unicidade de 96,0% das trajetórias, que se manteve antes e após a deduplicação e é visível nos mapas multiperspectiva (Figuras 1 e 2), é o achado com maior implicação para a gestão: existe um núcleo clínico reconhecível, porém a sequência, a frequência e os intervalos variam entre pacientes, o que sustenta a recomendação de um protocolo de navegação assistencial. A concentração de 89,9% do custo na quimioterapia é consistente com a literatura internacional de custo real em CCR [3].

A leitura descritiva da sobrevida mostra maior sobrevida no EOCRC (87,8% contra 62,7%); a ausência de significância estatística reflete o subdimensionamento amostral (poder de 15,4% e 8,9%), e não um resultado nulo confirmado. A idade contínua associou-se ao óbito na regressão de Cox, ao passo que a comparação categórica não alcançou significância estatística, o que é coerente com a perda de poder imposta pela dicotomização de uma variável contínua.

Embora conduzido com dados de uma operadora de saúde suplementar, o método é transferível ao Sistema

Único de Saúde (SUS). A abordagem apoia-se em logs de eventos administrativos, que o SUS também produz em larga escala nos sistemas de informação ambulatorial (SIA-SUS) e hospitalar (SIH-SUS) e nas autorizações de procedimentos de alta complexidade em oncologia (APAC), disponíveis no Departamento de Informática do SUS (DATASUS). A mesma sequência de mineração e verificação de conformidade poderia, assim, mapear a jornada assistencial do CCR na rede pública, subsidiar a avaliação de tecnologias em saúde e apoiar decisões de gestão, respeitadas as diferenças de codificação e de fluxo entre os sistemas.

As principais limitações são a ausência de TNM estruturado (contornada por *proxy*), o tamanho amostral do EOCRC (n=31), o custo baseado em autorizações e não em valor pago após glosas, e a resolução temporal de nível-dia. Recomenda-se a validação por painel de especialistas em duas frentes, os casos de Desvio Severo e os casos Indeterminados, e a reexecução completa na plataforma UpFlux para validação cruzada e *conformance checking* nativo.

5 CONCLUSÃO

A MP aplicada a 178 pacientes com CCR na saúde suplementar evidenciou jornada fragmentada (96,0% de trajetórias exclusivas), que se manteve após a correção do fracionamento de faturamento, com a quimioterapia concentrando 89,9% do custo e a idade contínua associada ao óbito. As diferenças descritivas favoráveis ao EOCRC em sobrevida e custo não foram confirmadas estatisticamente, limitação quantificada pelo poder do estudo e que não deve ser lida como ausência de efeito. Como próximo passo, recomenda-se a validação por painel de especialistas nos casos de Desvio Severo e Indeterminado, conferindo validade clínica aos achados automatizados antes de sua aplicação gerencial.

Contribuição dos autores. Os autores contribuíram igualmente em todas as etapas de construção do artigo: concepção e delineamento do estudo; coleta, análise e interpretação dos dados; redação e revisão crítica do manuscrito; e aprovação da versão final.

Agradecimentos. À Unimed Paraná pela autorização de acesso aos dados anonimizados. Estudo desenvolvido no PPGTS/PUCPR com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), no âmbito do programa Mestrado e Doutorado Acadêmico para Inovação (MAI/DAI), Processo nº 130038/2025-3, Chamada CNPq e Ministério da Ciência, Tecnologia e Inovações (MCTI) nº 009/2024. Conflito de interesse: a UpFlux é a ferramenta de mineração de processos utilizada pela operadora parceira; os autores declaram que a empresa não participou da concepção, análise, interpretação ou redação, e que nenhuma compensação foi recebida.

Financiamento. Os autores não declararam fonte de financiamento.

Disponibilidade de dados e materiais. Os autores não declararam a disponibilidade dos dados e materiais da pesquisa.

Conflitos de interesse. Os autores declaram que a UpFlux é a ferramenta de mineração de processos utilizada pela operadora parceira e que a empresa não participou da concepção, análise, interpretação ou redação; nenhuma compensação foi recebida.

Declaração de uso de IA. Os autores declaram o uso de ferramentas de IA generativa (Claude, da Anthropic) como apoio à revisão textual, à formatação do manuscrito e à geração de código para análise estatística. Toda a análise, interpretação e redação científica foram realizadas e validadas pelos autores. O uso foi declarado no formulário de submissão da plataforma Even3, conforme o edital do CBIS 2026.

REFERÊNCIAS

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3): 229-63. doi: https://doi.org/10.3322/caac.21834
2. Instituto Nacional de Câncer. Estimativa 2023: incidência de câncer no Brasil [Internet]. Rio de Janeiro: INCA; 2022 [citado 2026 nov 24]. Disponível em: https://docs.bvsalud.org/biblioref/2023/09/1510486/estimativa-2023.pdf
3. Jafari A, Hosseini FA, Jalali FS. A systematic review of the economic burden of colorectal cancer. Health Sci Rep. 2024;7(8):e70002. doi: https://doi.org/10.1002/hsr2.70002
4. Ugai T, Sasamoto N, Lee HY, Ando M, Song M, Tamimi RM, et al. Is early-onset cancer an emerging global epidemic? Current evidence and future implications. Nat Rev Clin Oncol. 2022;19(10):656-73. doi: https://doi.org/10.1038/s41571-022-00672-8
5. Murphy CC, Zaki TA. Changing epidemiology of colorectal cancer - birth cohort effects and emerging risk factors. Nat Rev Gastroenterol Hepatol. 2024;21(1):25-34. doi: https://doi.org/10.1038/s41575-023-00841-9
6. Puzzono M, Mannucci A, Grannò S, Zuppardo RA, Galli A, Danese S, et al. The Role of Diet and Lifestyle in Early-Onset Colorectal Cancer: A Systematic Review. Cancers (Basel). 2021;13(23):5933. doi: https://doi.org/10.3390/cancers13235933
7. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. CA Cancer J Clin. 2023;73(3):233-54. doi: https://doi.org/10.3322/caac.21772
8. Bretthauer M, Løberg M, Wieszczy P, Kalager M, Emilsson L, Garborg K, et al. Effect of Colonoscopy Screening on Risks of Colorectal Cancer and Related Death. N Engl J Med. 2022;387(17):1547-56. doi: https://doi.org/10.1056/NEJMoa2208375
9. van der Aalst W. Process Mining. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016. doi: https://doi.org/10.1007/978-3-662-49851-4
10. Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: A literature review. J Biomed Inform. 2016;61:224-36. doi: https://doi.org/10.1016/j.jbi.2016.04.007
11. Munoz-Gama J, Martin N, Fernandez-Llatas C, Johnson OA, Sepúlveda M, Helm E, et al. Process mining for healthcare: Characteristics and challenges. J Biomed Inform. 2022;127:103994. doi: https://doi.org/10.1016/j.jbi.2022.103994

12. Dallagassa MR, dos Santos Garcia C, Scalabrin EE, Ioshii SO, Carvalho DR. Opportunities and challenges for applying process mining in healthcare: a systematic mapping study. J Ambient Intell Human Comput. 2022;13(1):165-82. doi: https://doi.org/10.1007/s12652-021-02894-7
13. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-7. doi: https://doi.org/10.1016/S0140-6736(07)61602-X
14. van der Aalst W, Adriansyah A, de Medeiros AKA, Arcieri F, Baier T, Blickle T, et al. Process Mining Manifesto. In: Lecture Notes in Business Information Processing. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 169-94. doi: https://doi.org/10.1007/978-3-642-28108-2_19
15. Sociedade Brasileira de Oncologia Clínica. Diretrizes de tratamento do câncer colorretal. São Paulo: SBOC; 2026.
16. Kaplan L, Meier P. Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association. 1958;53(282):457-81. doi: https://doi.org/10.1080/01621459.1958.10501452
17. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1972;34(2):187-202. doi: https://doi.org/10.1111/j.2517-6161.1972.tb00899.x
18. Schoenfeld DA. Sample-Size Formula for the Proportional-Hazards Regression Model. Biometrics. 1983;39(2):499. doi: https://doi.org/10.2307/2531021
19. Instituto Brasileiro de Geografia e Estatística. Índice Nacional de Preços ao Consumidor Amplo (IPCA) [Internet]. Rio de Janeiro: IBGE; 2024 [citado 2026 out 18]. Disponível em: https://apisidra.ibge.gov.br/values/t/1737/n1/all/v/2266/p/202412
20. Davidson-Pilon C. lifelines: survival analysis in Python. JOSS. 2019;4(40):1317. doi: https://doi.org/10.21105/joss.01317

NOTAS EDITORIAIS

O Journal of Health Informatics é uma publicação científica mantida pela Sociedade Brasileira de Informática em Saúde, associação científica sem fins lucrativos dedicada ao desenvolvimento da Informática em Saúde e da Saúde Digital no Brasil. O processo editorial é conduzido com independência científica, sem interferência institucional, comercial ou de patrocinadores, e segue as boas práticas internacionais de ética em publicação científica.

Este artigo foi avaliado por pares, em sistema duplo-cego, por especialistas da área. A equipe editorial usou ferramentas de inteligência artificial, com supervisão humana, na triagem inicial, na conferência do texto, das referências e dos números e, quando necessário, no redesenho de figuras, quadros e tabelas, aprovado pelos autores na prova. As ferramentas e as salvaguardas estão na Política de uso de inteligência artificial da revista. As decisões editoriais foram tomadas por pessoas.

SBIS Sociedade Brasileira de Informática em Saúde