Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cancer patient pathways and care coordination in Kuwait: an exploratory process mining study using linked health data

Munirah Alateeqi
Kuwait University
Abdulwahab Altourah
Kuwait Cancer Control Center
Dari Alhuwail
Kuwait University
Jillian Timperley
Creighton University School of Medicine
Waddah Al-Refaie
Creighton University School of Medicine
Abdullah Alibrahim
abdullah.alibrahim@ku.edu.kw

Kuwait University

Research Article

Keywords:

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11310775/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Cancer patient pathways and care coordination in Kuwait: an exploratory process mining study using linked health data

Munirah Alateeqi[1]; Abdulwahab Altourah[2]; Dari Alhuwail[3,6,7]; Jillian Timperley[4]; Waddah Al-Refaie[4]; Abdullah Alibrahim[1,4,5,6*]

[1] Department of Industrial & Management Systems Engineering, College of Engineering & Petroleum, Kuwait University, Kuwait City, Kuwait

[2] Kuwait Cancer Control Center, Ministry of Health, Kuwait City, Kuwait

[3] Information Science Department, College of Life Sciences, Kuwait University, Kuwait City, Kuwait

[4] Creighton University School of Medicine, Omaha, USA

[5] Middle East Initiative, Harvard Kennedy School, Cambridge, MA, USA

[6] GeoHealth Lab, Dasman Diabetes Institute, Kuwait City, Kuwait

[7] Communication and Information Technology Regulatory Authority, Kuwait City, Kuwait

[*] Corresponding author: abdullah.alibrahim@ku.edu.kw

Abstract

Background: The rising incidence and complexity of cancer treatment increase the need to understand care delivered across specialties and facilities. Evidence on cancer pathways remains limited in settings with fragmented health data. This study evaluated the feasibility of linking multiple healthcare data sources for process mining and described breast and colorectal cancer pathways in Kuwait.

Methods: In this retrospective cohort study, we linked two administrative datasets: patient records from the Kuwait Cancer Control Center (KCCC) Electronic Health Record (EHR) system covering 2015 through the first quarter of 2023, and national hospital discharge data from the Ministry of Health (MOH) for 2015, 2016, 2019 and 2020. Fifty patients were purposively selected, 25 with breast cancer and 25 with colorectal cancer. The linked event log comprised 3,334 activity records and was examined using process discovery, event classification and Sankey visualization.

Results: Colorectal cancer patients recorded more clinic visits per week than breast cancer patients (0.31 vs. 0.23, $p<0.05$) and more diagnostic events per patient (26.0 vs. 12.2, $p<0.05$),

driven by laboratory and radiology testing. Breast cancer patients recorded more radiation oncology activity ($p<0.05$). Average captured journey length did not differ (51.4 vs. 49.5 months, $p = 0.726$). Medical oncology appeared in 92% of breast and 100% of colorectal cancer journeys. General hospitals accounted for 24 of 101 (24%) colorectal hospitalizations, compared with 8 of 74 (11%) breast cancer hospitalizations.

Conclusions: Linking MOH discharge data with KCCC EHR records was feasible and made multi-facility cancer trajectories visible. The observed pathway differences are hypothesis-generating and identify priorities for larger, representative studies and for more complete, standardized event capture.

Keywords: Process mining; Cancer care pathways; Care coordination; Electronic health records; Health information systems; Kuwait

1 Background

An estimated 18.5 million new cancer cases, excluding non-melanoma skin cancers, occurred globally in 2023 [1]. Improved survival and increasingly multimodal treatment have also made cancer care more longitudinal and complex, requiring patients to move among oncologists, surgeons, radiologists, nurses, and other services [2–7]. Healthcare operations management literature has increasingly recognized care coordination as a fundamental organizational design challenge, where the deliberate organization of patient care activities between multiple participants requires purposeful analytical approaches to understand actual coordination patterns versus prescribed clinical pathways [8, 9]. Cancer care delivery exemplifies multi-team systems coordination challenges, where specialized teams must coordinate comprehensive care across organizational boundaries while maintaining quality, efficiency, and patient-centered outcomes [10].

Traditional healthcare operations research has often focused on resource allocation and patient flow optimization [11]. Analyses based only on prescribed pathways may not capture how patients actually move through services, potentially obscuring variation and gaps in recorded care [8, 9, 12]. Process mining can reconstruct recorded pathways from electronic health record (EHR) and administrative data and describe how care is distributed across services [13, 14]. Kuwait's Vision 2035 initiatives

emphasize standardized and coordinated cancer care, supported by infrastructure investment [15], and the Kuwait Cancer Control Center's (KCCC) strategic plan targets streamlined patient pathways and stronger departmental collaboration [16]. At the same time, fragmented records and inconsistent documentation constrain systematic analysis of care delivery in Kuwait and in other health systems with similar data limitations [17–19].

Cancer care coordination analysis requires innovative data integration approaches because treatment spans multiple facilities, from specialized cancer centers to general hospitals delivering supportive care, making single-source datasets inherently incomplete. Administrative data, generated across healthcare encounters, offer an opportunity to more comprehensively capture the cancer care journey. Prior work in lung cancer clinical pathways demonstrates that multi-source data linkage can improve pathway completeness beyond what individual institutions can achieve [20]. Nevertheless, linking heterogeneous healthcare databases presents substantial challenges, including data inconsistencies, missing patient identifiers, and linkage errors that may compromise analytical validity [21].

Cancer care in Kuwait has not been mapped using linked cancer-center EHR and national discharge data. We therefore evaluated whether these sources could be integrated into a process-mining event log and used to describe recorded breast and colorectal cancer pathways, cross-specialty activity patterns, and multi-facility utilization. We also examined the data-quality constraints that affect pathway reconstruction. The analysis was exploratory and intended to generate hypotheses for larger, representative studies in Kuwait and in health systems facing similar data-fragmentation challenges.

2 Methods

2.1 Study Design

This exploratory retrospective cohort study evaluated the feasibility of integrating multiple administrative health data sources for process mining of cancer patient pathways. It prioritized methodological demonstration and descriptive pathway characterization over statistical generalization. Reporting follows the STROBE statement for observational studies; the completed checklist is

provided as an additional file. An LLM-based tool was used for language editing and consistency checks only; it was not used for study design, data analysis, interpretation, or figure creation.

2.2 Data Sources

This study utilized two administrative healthcare datasets from Kuwait's healthcare system. The primary dataset consisted of records from the EHR system from the KCCC, encompassing comprehensive patient encounter metadata across multiple specialized oncology clinics from the inception of the electronic health record system in 2015 through the first quarter of 2023. KCCC is Kuwait's primary specialized cancer treatment facility, providing comprehensive oncology services, including medical oncology, surgical oncology, radiation oncology, and supportive care.

The second dataset comprised national-level hospital discharge data from the Ministry of Health (MOH), covering discharge summaries from 19 public hospitals coded using the International Classification of Diseases, 10^{th} Revision (ICD-10) classification system. MOH data were available for four selected years (2015, 2016, 2019, 2020), representing routine administrative data collected for national healthcare reporting and monitoring purposes. These datasets were selected to capture both specialized cancer care delivery at Kuwait's primary oncology center and broader healthcare utilization across the national public hospital system, enabling comprehensive analysis of cancer patient pathways spanning multiple healthcare facilities and care settings.

Throughout this paper, hospitals are grouped into three categories: KCCC, the national specialized cancer center; *specialized* hospitals, which provide dedicated services in a single clinical domain other than oncology; and *general* (secondary) hospitals, which provide general acute inpatient care. Hospitalizations recorded at any facility other than KCCC are described as care delivered outside the cancer center.

2.3 Study Sample and Rationale

The sampling frame was the MOH 2019–2020 national discharge dataset. Patients were eligible if they (1) were referred to KCCC in 2019, which ensured they would also be captured in the KCCC electronic health record, (2) had at least two admissions to MOH hospitals during the study period,

and (3) carried a primary cancer diagnosis coded C50 for breast cancer or C18–C20 for cancer of the colon or rectum. Applying these criteria yielded 120 eligible breast cancer patients and 54 eligible colorectal cancer patients, from which 25 of each were purposively selected for analysis. KCCC electronic health record data were then retrieved for the 50 selected patients, and no selected patient was subsequently excluded. This sample size was chosen to demonstrate the feasibility of data integration and process mining methodology while acknowledging limitations for statistical generalization. Table 1 summarizes the hospitalizations and hospitals represented in the MOH discharge data for each cohort.

Because the study used retrospective administrative records, there was no active follow-up or patient contact. Each observation window ran from the first activity recorded in either source to the last. Care delivered outside the available sources or periods was not observed, and absence of a recorded event was not interpreted as absence of care.

Several design features were intended to reduce measurement differences between cohorts. Drawing the frame from national discharge data and requiring referral to KCCC reduced the single-source ascertainment gap that would otherwise truncate multi-facility pathways, at the cost of excluding patients treated within only one part of the system. Requiring at least two MOH admissions ensured pathways rich enough to test linkage and process mining, but necessarily selected patients with more inpatient contact than the wider cancer population. A consultant medical oncologist assigned the order of ambiguous same-day activity sequences on clinical grounds before analysis. Both cohorts were drawn from the same two sources and periods and processed using the same activity definitions and event-log pipeline.

Table 1: Hospitalizations and hospitals represented in the MOH national discharge data, by cancer type

Data Set	**Breast Cancer**	**Colorectal Cancer**
MOH 2019–2020 national-level discharge data	69 hospitalizations; 5 general hospitals; 2 specialized hospitals	96 hospitalizations; 4 general hospitals; 2 specialized hospitals
MOH 2015–2016 national-level discharge data	6 hospitalizations; 1 general hospital; 1 specialized hospital	7 hospitalizations; 2 general hospitals; 1 specialized hospital

Table 2: Characteristics of the study cohort at registration

Characteristic	**Breast cancer (n = 25)**	**Colorectal cancer (n = 25)**
Age at registration, years, mean (SD)	50.4 (11.7)	51.2 (12.7)
Age at registration, years, median (range)	50 (24–72)	51 (30–74)
Female sex, n/N with data (%)	25/25 (100)	7/11 (64)
Kuwaiti nationality, n/N with data (%)	2/13 (15)	3/11 (27)
Primary site recorded at diagnosis, n		
Breast, subsite not specified	19	—
Breast, named quadrant or central portion	4	—
Colon (unspecified, sigmoid, descending)	—	16
Rectum or rectosigmoid junction	—	7
Other site recorded in the diagnosis field	1	2
No diagnosis record	1	0
Observation window, months, mean (95% CI)	51.4 (44.7, 58.0)	49.5 (41.6, 57.5)
Patients with missing data from EHR, n		
Sex	0	14
Nationality	12	14
Primary site	1	0

Cohort characteristics are summarized in Table 2. Age at registration was derived from the date of birth encoded in the national civil identification number. Nationality was obtained by linking the civil identification number to the KCCC demographic register; 24 of the 50 patients were present in that register, so nationality is reported over the subset with available data and the number of patients missing each variable is given. All 25 breast cancer patients were female. The two cohorts were of similar age ($p = 0.80$). Sex is collinear with cancer type in this sample, since the entire breast cancer cohort is female while the colorectal cohort is mixed, so the cohorts should not be read as differing only in cancer site. A small number of records list a diagnosis site outside the sampling criteria. These may represent secondary or metastatic sites recorded in the same field, but their clinical status could not be verified from the available extract.

A total of 18 activities in the dataset were grouped into 5 categories: administrative, diagnostic, inpatient, outpatient, and treatment. These were structured by location, activity name, and classification. The standardization process ensured consistency and enabled analysis across both KCCC and MOH datasets.

Breast and colorectal cancers were selected because they are among the most prevalent cancer types in Kuwait and provide clinically distinct pathways with sufficiently rich event histories to test data

linkage and process mining.

2.4 Key Process Measures

Patient pathways were operationalized as sequential chains of healthcare activities experienced by individual patients across time and treatment facilities, measured through process mining event logs containing patient identifiers, activity types, timestamps, and location attributes [22]. Each pathway encompassed the complete captured journey from initial registration through ongoing care, with pathway characteristics quantified using journey length (months), event frequency (activities per patient), activity intensity (visits per week), and transition patterns between care activities.

Care coordination and team interactions were not directly observed. We instead used three descriptive pathway proxies: (1) recorded transitions between specialties and departments, which do not necessarily represent formal referrals; (2) the frequency and patient coverage of recurring services, particularly medical oncology; and (3) elapsed time between consecutive recorded activities. These measures describe the structure of recorded care and should not be interpreted as direct measures of collaboration quality or clinical appropriateness.

2.5 Data Linking

For data linking, a systematic process was followed, starting with the cleaning and standardization of patient records. Repeated same-day activities for the same patient were collapsed, and date formats were unified. Deterministic linkage using national ID, date of birth, sex, and file number was conducted in Microsoft Excel, first within MOH datasets, and then between MOH and KCCC datasets. Matching records were merged into a single linked dataset. In this linking method, a pair of records was treated as a match only if all of the chosen identifiers matched exactly [23].

2.6 Data Preparation

A common challenge in preparing raw data for process mining is imprecise activity timing, especially for same-day activities at KCCC. A consultant medical oncologist with 20 years' experience in cancer care reviewed ambiguous same-day combinations and specified their most plausible clinical

order based on established protocols and observed care patterns. This ordering rule was then applied consistently to assign within-day sequences. These assigned times support event ordering but are not observed appointment times; sub-day intervals involving them should therefore not be interpreted as measured waiting times. Suriadi et al. describe related approaches to event-time imperfections in process mining [24].

Event logs are the primary input for process mining algorithms [25]. A minimum event log consists of a case identifier, activity, timestamp, and optional activity attributes [22]. Each component and its corresponding variable is shown in Table 3. Process discovery was implemented using Disco and R Statistical Software with the bupaR package [26–28]. Activity aggregation, linkage, and within-day ordering reduced artificial inflation of pathway complexity and provided a consistent basis for estimating service touchpoints and transition frequencies.

Table 3: Event log components

Event log component	**Explanation**	**The corresponding variable**
Case ID	Study-specific patient identifier	Pseudonymous case ID
Activity	Action in the process	Activities
Timestamp	The time assigned to the event	Recorded or assigned activity time
Other attributes	Any additional case or activity attributes	Activity classification

2.7 Statistical Analysis

Descriptive statistics were calculated for pathway characteristics including observation-window length, event frequency, and activity patterns. The patient was the unit of analysis for between-group comparisons. Breast and colorectal cancer cohorts were compared using two-sided Welch two-sample t-tests, which do not assume equal variances, and 95% confidence intervals for group means were derived from the t distribution. Statistical significance was set at $p<0.05$. Because the analysis was exploratory and included many activity-level comparisons, no adjustment was made for multiplicity; results with $0.05 \leq p < 0.10$ were treated as suggestive only. Reported p-values should therefore be read as descriptive rather than confirmatory. All estimates were unadjusted; no confounders were controlled for, and no subgroup or interaction analyses were conducted beyond

the planned comparison of the two cohorts. Ambiguous same-day timestamps were handled using the clinically reviewed ordering procedure described above. Missing demographic variables were not imputed. No formal sensitivity analyses were performed, although process-map abstraction levels were varied during exploratory mapping. Analysis was performed using R version 4.1.2 [28].

2.8 Exploratory Process Mapping

The exploratory analysis had two phases: log and pattern inspection, followed by process discovery. The first phase summarized the event logs, key process variables, and event distributions, with confidence intervals for group means.

In the discovery phase, simplified process maps were created in Disco using its implementation of Fuzzy Miner [29], which reduces complex models using significance and correlation metrics [30]. Sankey diagrams visualized transitions around two high-frequency activities: day care for breast cancer and laboratory testing for colorectal cancer. The diagrams show recorded activity flows and mean elapsed times, not formal referrals or the quality of coordination.

2.9 Ethical Considerations

This study was approved by the Ministry of Health Institutional Review Board (IRB Approval No. #1594, 1644/2021, renewed in 2025). Records were linked on national identifiers and then de-identified prior to analysis, and all procedures adhered to local data protection regulations.

3 Results

3.1 Data Linking and Validation Results

All low-level laboratory activities recorded for the same patient on the same day were combined into a single daily entry. Deterministic linkage produced exact agreement on the selected identifiers for the included cohort, and no ambiguous matches were observed, supporting the technical feasibility of integrating the two sources. Within-day ordering was assigned 194 times across 48 unique activity

combinations in the breast cancer dataset and 262 times across 50 combinations in the colorectal cancer dataset. The hospitalization counts in Table 1 exceed those carried into the linked event log by one admission for breast cancer and two for colorectal cancer. These episodes crossed an MOH reporting-year boundary and appeared in annual source extracts without resolving to a distinct hospitalization event within the linked observation window.

After aggregating semantically equivalent records into unified activities (for example, multiple laboratory orders collapsed into a single "Lab" activity per patient-day), the analysis event log contained 1,435 events for breast cancer and 1,899 events for colorectal cancer, giving 3,334 events in total. All process maps and summary statistics reported below are computed on this log.

3.2 Exploratory Process Mining of Cancer Patient Administrative Data

After the data-handling phase, process mining techniques were applied to the linked datasets to study cancer patient treatment journeys. The data were prepared as event logs for import into Disco and R for analysis [26, 28]. An overview of the cancer care process was provided by summarizing an event log, which recorded 3,334 events; colorectal cancer patients accounted for 57.0% of these events.

Colorectal cancer patients had slightly more activity types (16 vs. 14) across more hospitals (9 vs. 8), and all 50 patients had distinct recorded paths. Breast cancer patients had a slightly longer mean observation window (51.4 vs. 49.5 months, $p = 0.726$), whereas colorectal cancer patients had more events on average (76.0 vs. 57.4, $p = 0.115$) and more clinic days per week (0.31 vs. 0.23, $p<0.05$). Table 4 summarizes the key variables; activity coverage is shown in Figures 3(b) and 4(b).

Table 4: Study key variables

Variable	**Breast Cancer** Mean (95% CI)	**Colorectal Cancer** Mean (95% CI)
Mean observation-window length (months)	51.4 (44.7, 58.0)	49.5 (41.6, 57.5)
Mean number of events per patient	57.4 (46.5, 68.2)	76.0 (56.1, 95.7)
Mean number of clinic days per week	0.23 (0.19, 0.26)	0.31 (0.24, 0.38)

Outpatient visits were the most frequent event class (46% overall; breast: 47% [27.2 events per patient], colorectal: 44% [33.8], $p = 0.236$). Diagnostic events were more frequent among colorectal cancer patients (29% overall; breast: 21% [12.2 events per patient], colorectal: 34% [26.0], $p<0.05$). Treatment events were more frequent among breast cancer patients (19% overall; breast: 24% [14.0 events per patient], colorectal: 15% [11.2], $p = 0.328$). Inpatient events accounted for 5% of events overall and were more frequent per patient in the colorectal cancer cohort (breast: 3.0, colorectal: 4.0, $p = 0.049$); administrative events were least frequent. Of 74 breast cancer hospitalizations, 63 (85%) occurred at KCCC, 8 (11%) at general hospitals and 3 (4%) at other specialized hospitals. Of 101 colorectal cancer hospitalizations, 73 (72%) occurred at KCCC, 24 (24%) at general hospitals and 4 (4%) at other specialized hospitals. Figure 1 shows group means and confidence intervals by event class.

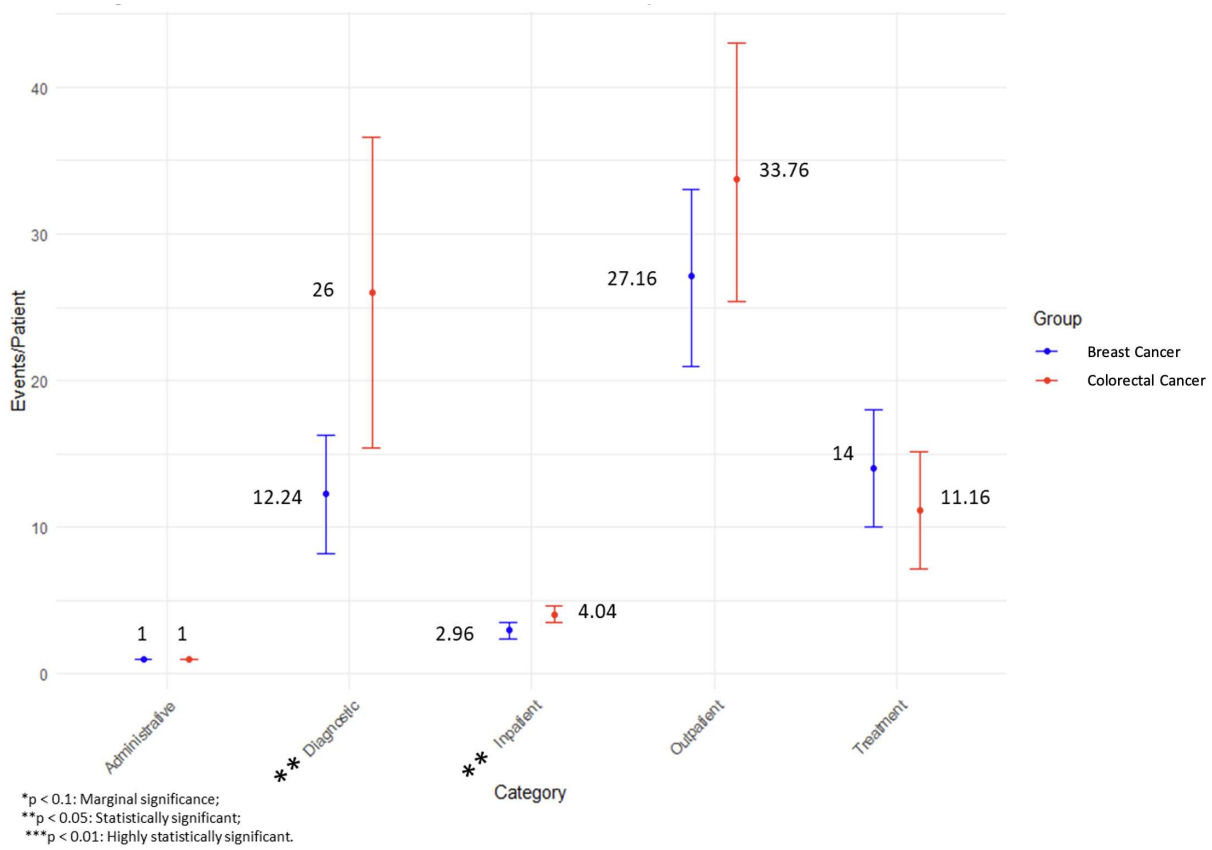


Figure 1: Event classes by cancer type. Points show mean events per patient; bars show 95% confidence intervals.

Figure 2 shows events per patient by activity. Laboratory events were more frequent among colorectal

cancer patients (p<0.05), while the radiology difference was suggestive ($p = 0.088$). Medical oncology was the most frequent outpatient activity in both groups, with no evidence of a difference ($p = 0.377$). Surgical oncology event counts were nearly identical (p>0.99), and anesthesia/ICU event counts did not differ ($p = 0.155$). Radiation oncology visits were more frequent among breast cancer patients (p<0.05), while the difference in pain clinic visits was suggestive ($p = 0.063$). For treatment activities, radiation oncology was more frequent among breast cancer patients (p<0.05); the difference in day care treatments was suggestive ($p = 0.089$). Day care refers to same-day therapies in the hospital infusion room.

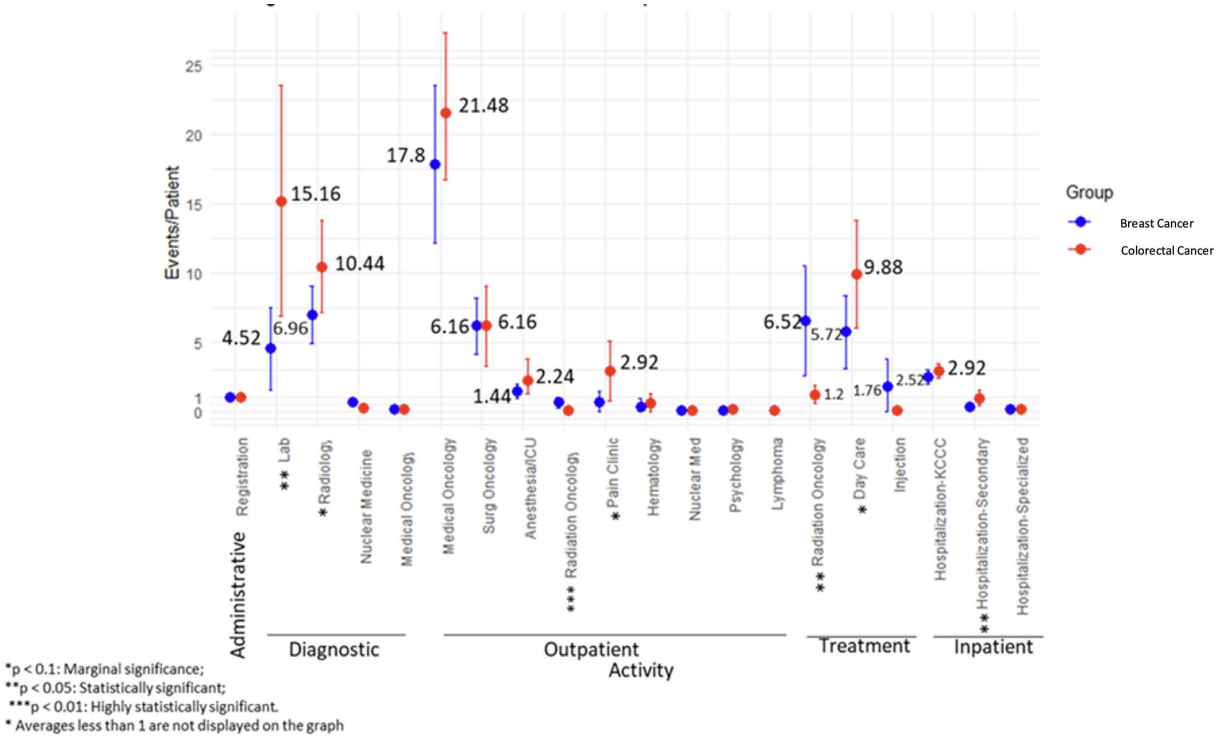


Figure 2: Activities by cancer type. Points show mean events per patient; bars show 95% confidence intervals.

3.3 Process Mapping of Cancer Patient Journeys

Simplified process maps were created in Disco using adjusted abstraction levels (85% activities and 0% paths for breast cancer; 75% activities and 0% paths for colorectal cancer). The thresholds were chosen during exploratory mapping to display 12 activities in each cohort and improve legibility; the resulting maps are descriptive and should not be used for direct quantitative comparison of

omitted paths. Thicker arrows indicate more frequent displayed transitions, while darker activity colors indicate greater frequency, coverage, or mean elapsed time, depending on the panel.

Figure 3(a) shows the breast cancer frequency map. Medical oncology was the most frequent recurring activity (448 occurrences), followed by radiation oncology (180), radiology (174), surgical oncology (154), day care (143), laboratory testing (113), hospitalization (74), and the injection room (44). Pain clinic, nuclear medicine, and anesthesia/ICU activities occurred less often.

Figure 3(b) shows patient coverage in the displayed activities. Registration and hospitalization each covered 100% of patients, medical oncology 92%, radiology 96%, radiation oncology 88%, day care 68%, and pain clinic 16%.

Figure 3(c) shows mean elapsed times between displayed activities. Registration to medical oncology averaged 15.1 weeks and included an outlier exceeding 1 year. Hospitalization to laboratory testing had the longest mean interval (49.4 weeks). Consecutive medical oncology events averaged 44.2 days, surgical to radiation oncology 18.3 days, registration to surgical oncology 3.3 days, day care to hospitalization 41.2 hours, and radiology to anesthesia/ICU 21.7 hours. Sub-day values can reflect the clinically assigned within-day order described in Methods.

For colorectal cancer patients, Figure 4(a) shows that medical oncology was the most frequent recurring activity (541 occurrences), followed by laboratory testing (379), radiology (261), day care (247), surgical oncology (154), hospitalization (101), and pain clinic (73). Radiation oncology (31), hematology (16), and nuclear medicine (8) were less frequent.

Figure 4(b) shows patient coverage. Registration, medical oncology, day care, and hospitalization each covered 100% of patients; surgical oncology, radiology, and laboratory testing each covered 96%; and anesthesia/ICU covered 80%. Coverage was 56% for radiation oncology, 44% for pain clinic, 24% for hematology, and 20% for nuclear medicine.

Figure 4(c) shows mean elapsed times. Registration to radiation oncology averaged 14.1 weeks and included an outlier of approximately 1 year. Other longer displayed intervals were hospitalization to registration (17.5 weeks), hospitalization to laboratory testing (25.7 weeks), and radiation oncology to surgical oncology (25.9 weeks). Consecutive medical oncology events averaged 36.7 days, pain clinic to laboratory testing 18.2 days, and day care to hematology 5.5 days.

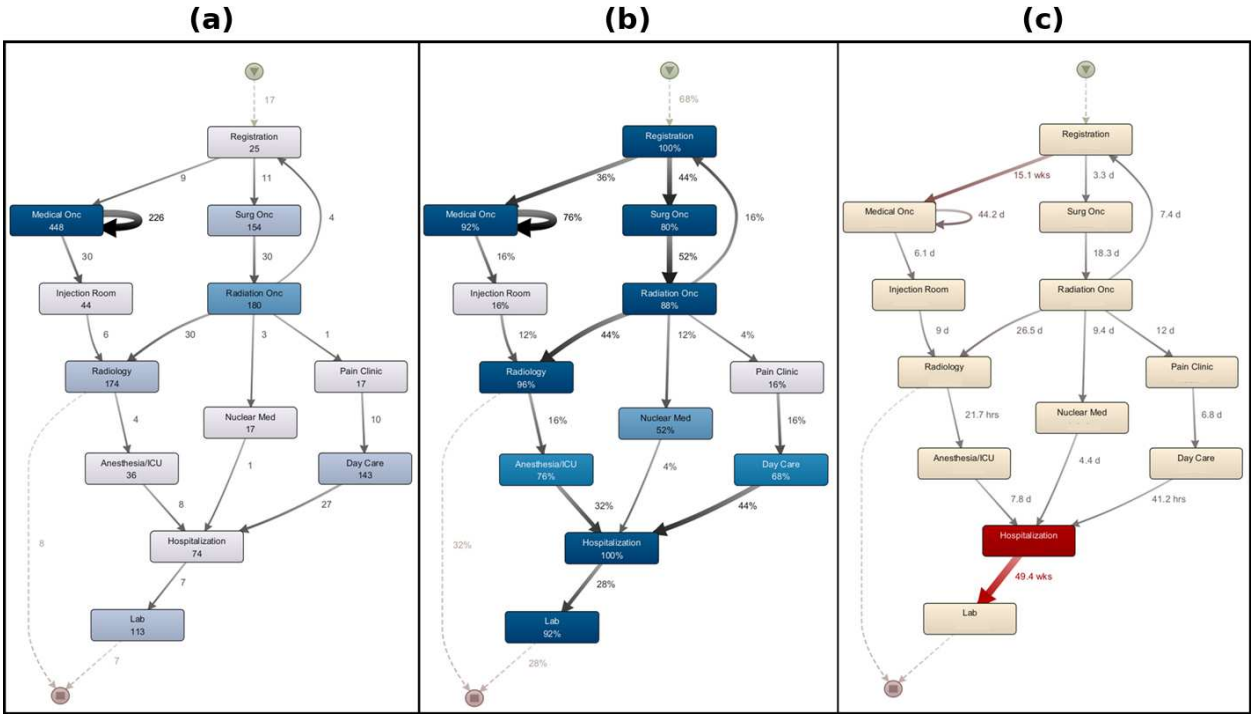


Figure 3: Breast cancer process maps. (a) Activity frequency; (b) patient coverage; (c) mean elapsed times.

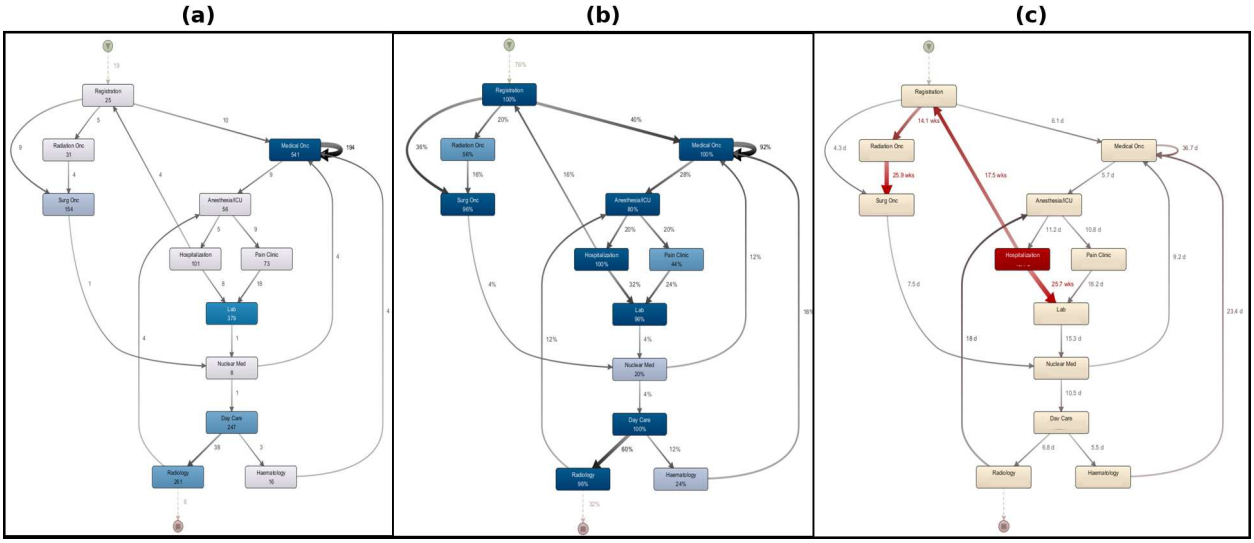


Figure 4: Colorectal cancer process maps. (a) Activity frequency; (b) patient coverage; (c) mean elapsed times.

Sankey diagrams were developed around day care for breast cancer and laboratory testing for colorectal cancer. Percentages represent shares of observed inbound or outbound transitions, not shares of unique patients; durations are mean elapsed times between consecutive recorded activities.

Figure 5 shows transitions around day care for breast cancer patients. Of inbound transitions, 57% came from medical oncology (mean interval 16.3 days), 13% from radiology (6.3 days), 10% from pain clinic (6.8 days), and 10% from hospitalization (11.6 days). Of outbound transitions, 29% went to medical oncology (18.5 days), 28% returned to day care (17.4 days), and 18% went to hospitalization (2 days); radiology, laboratory testing, and pain clinic each accounted for less than 10%.

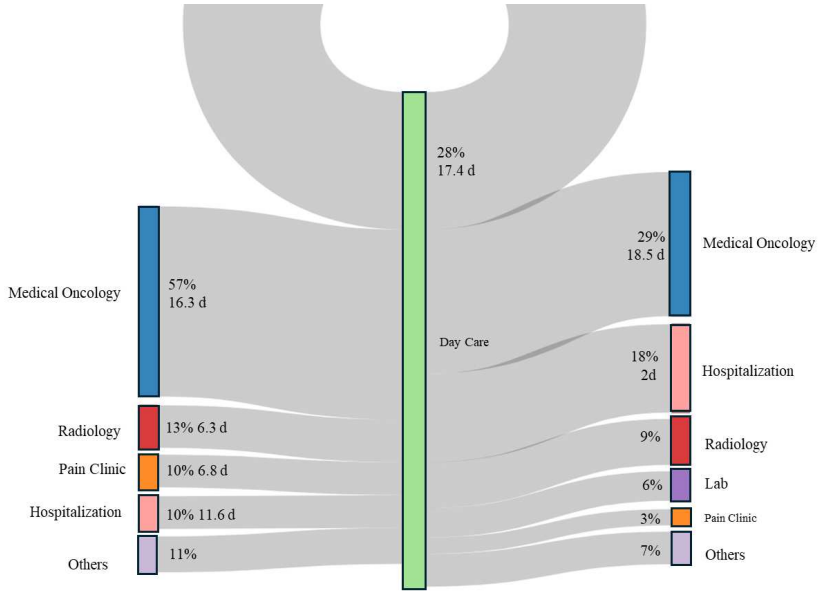


Figure 5: Breast cancer transitions around day care. Percentages are transition shares; self-transitions appear only among outflows.

Figure 6 shows transitions around laboratory testing for colorectal cancer patients. Of inbound transitions, 34% came from medical oncology (mean interval 31 days), 33% from radiology (32.6 days), and 16% from day care (13.5 days). Of outbound transitions, 25% went to medical oncology (13 days), 20% to radiology (15.9 days), and 6% to day care. Laboratory self-transitions accounted for 45% of outbound transitions and averaged 11.4 days.

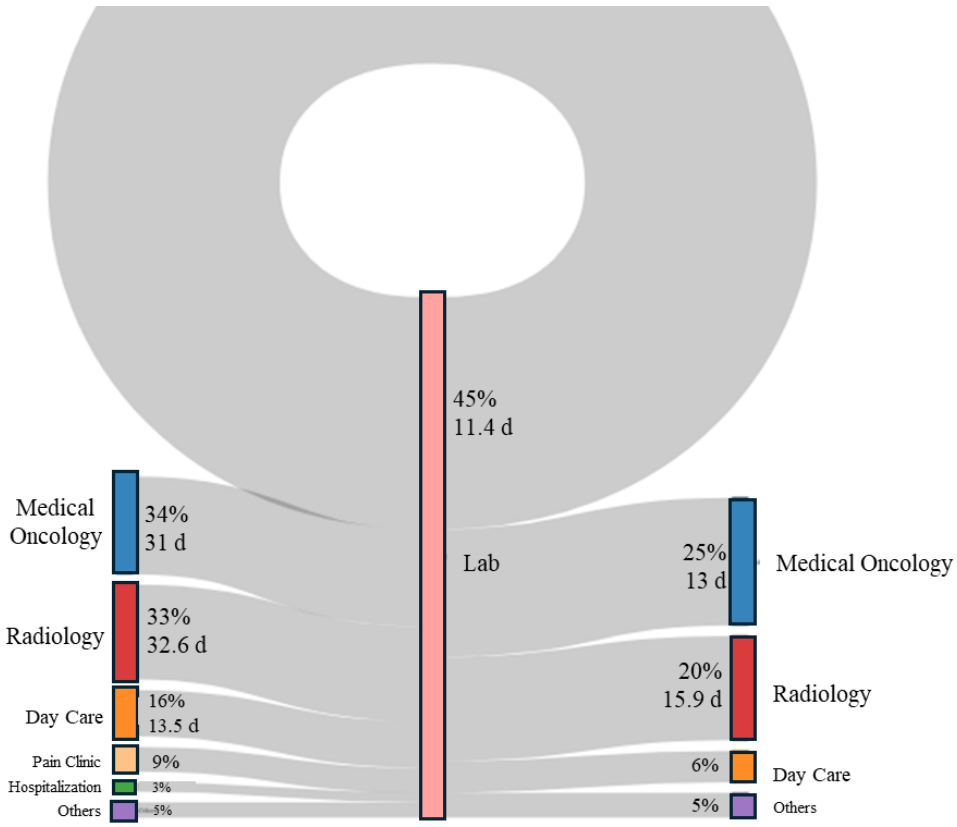


Figure 6: Colorectal cancer transitions around laboratory testing. Percentages are transition shares; self-transitions appear only among outflows.

4 Discussion

This study examined recorded pathways for 50 purposively selected patients, 25 with breast cancer and 25 with colorectal cancer, using linked MOH discharge and KCCC EHR data from 2015 to 2023. The analysis demonstrates the technical feasibility of multi-source process mining in a fragmented data environment and describes pathway complexity, visit intensity, and multi-facility utilization in this selected cohort. It also identifies data-quality limitations, including imprecise timestamps and incomplete capture, that affect the reliability of pathway measurement.

In this cohort, colorectal cancer patients had more recorded events across a broader set of hospitals. Although mean observation-window length did not differ, colorectal cancer patients had more clinic days per week. Outpatient visits were the most frequent event class in both groups, consistent with the expansion of ambulatory cancer care documented in other health systems [31].

Colorectal cancer patients had more diagnostic activity, particularly laboratory and radiology events, whereas breast cancer patients had more radiation oncology activity. Colorectal cancer patients also had more recorded inpatient events and a larger share of hospitalizations in general hospitals outside KCCC. These differences may reflect variation in clinical need, disease stage, sex composition, or care capture; the available data do not distinguish among these explanations.

Process maps showed that medical oncology was a frequent recurring activity in both cohorts, alongside transitions involving surgical oncology, radiation oncology, and diagnostic services. Coverage maps identified services that appeared in most recorded pathways, while the duration maps showed substantial variability and sensitivity to outliers and assigned within-day ordering. The Sankey diagrams localized high-volume day care and laboratory cycles. These patterns identify hypotheses for future evaluation of scheduling, pathway standardization, and diagnostic throughput; they do not by themselves establish coordination quality or inefficiency.

Collectively, the findings show greater recorded diagnostic, inpatient, and clinic activity in the colorectal cancer cohort and greater radiation oncology activity in the breast cancer cohort, with no evidence of a difference in observation-window length. Improved data standardization and broader source coverage would make these pathway comparisons more reliable.

From a health systems operations perspective, this study illustrates the types of measures that linked administrative data can provide, including clinic-day intensity, activity coverage, and multi-facility utilization. The difference between 0.31 and 0.23 clinic days per week shows how pathway data could inform demand forecasting once estimated in a larger, representative population; the present purposive sample is not suitable for staffing or capacity estimates.

Medical oncology appeared in 92–100% of recorded pathways, suggesting that future work should examine its role in cross-specialty coordination using direct referral and communication data. The 456 assigned within-day sequences (194 breast and 262 colorectal) also identify a concrete information-system priority: capture of reliable event times and provenance.

4.1 Limitations and Mitigation Strategies

This study should be interpreted in light of several limitations. First, the small purposive sample limits power and generalizability; the results demonstrate feasibility and generate hypotheses rather than population estimates. Second, requiring KCCC referral and at least two MOH admissions preferentially selected patients with repeated inpatient contact and may overrepresent pathway complexity and multi-facility use. Conversely, care outside the linked sources may understate total service use. Third, many activity-level comparisons were performed on 25 patients per group without multiplicity adjustment. Welch t-tests were applied to event counts that may be skewed, so p-values and confidence intervals should be interpreted descriptively. Fourth, the cohorts differed in sex composition, sex was missing for 14 colorectal cancer patients, and nationality was available for only 24 of 50 patients. No adjustment for patient characteristics was possible. Stage at diagnosis was unavailable, preventing assessment of the most important likely clinical confounder. Fifth, same-day activity order was assigned by clinical rules rather than observed timestamps; consequently, sub-day intervals are not measured waiting times, and no sensitivity analysis assessed alternative orderings. Sixth, process mining describes recorded transitions but does not establish referrals, collaboration quality, clinical appropriateness, or causal relationships. Seventh, MOH discharge data were available only for 2015, 2016, 2019 and 2020, leaving intervening hospitalizations uncaptured. Eighth, care in the private sector, other outpatient settings, or abroad may create silent intervals and truncated pathways. The findings are therefore specific to the observed KCCC and MOH data

footprint.

4.2 Implications

This work demonstrates a process for linking routine data and reconstructing recorded cancer pathways in a setting with fragmented information systems. With larger representative samples and more complete timestamps, the same approach could support health-services researchers and operations managers in describing patient flow, identifying cross-facility activity, and prioritizing questions for quality improvement.

Future research should use larger cohorts, continuous source coverage, direct referral or communication data, and conformance analyses to distinguish expected clinical variation from potentially avoidable delay or fragmentation. Evaluating links between pathway patterns and clinical or patient-reported outcomes would also be necessary before using these measures to guide care redesign.

5 Conclusions

Linking KCCC EHR and national discharge data for process mining was feasible and revealed pathway information unavailable from either source alone. In this selected cohort, colorectal cancer patients had greater recorded clinic-day intensity and broader facility use, while medical oncology appeared in nearly all pathways. These findings are hypothesis-generating. Larger representative studies with more complete event times, clinical covariates, and cross-sector data are needed before the measures can support capacity planning or care redesign.

List of abbreviations

CI: Confidence interval; EHR: Electronic Health Record; ICD-10: International Classification of Diseases, Tenth Revision; ICU: Intensive Care Unit; IRB: Institutional Review Board; KCCC: Kuwait Cancer Control Center; KFAS: Kuwait Foundation for the Advancement of Sciences; MOH: Ministry of Health; STROBE: STrengthening the Reporting of OBservational studies in Epidemiology.

Declarations

Ethics approval and consent to participate

Ethics approval for this retrospective study using linked administrative and EHR data was obtained from the Ministry of Health Institutional Review Board (IRB) in Kuwait (approval #1594, 1644/2021, renewed in 2025). Consent to participate was waived for the retrospective analysis of routinely collected administrative and EHR data, as approved by the ethics committee. Records were linked on national identifiers and then de-identified prior to analysis, and all procedures adhered to local data protection regulations.

Consent for publication

Not applicable. No individually identifiable patient data are reported.

Availability of data and materials

The datasets analyzed during the current study are not publicly available due to privacy restrictions related to patient health information governed by the Kuwait Ministry of Health. Any request for access would be subject to approval by the Ministry of Health and the relevant institutional review board, and enquiries may be directed to the corresponding author. Analytic code and processing scripts may likewise be made available from the corresponding author on reasonable request, subject to institutional review and any applicable data-governance constraints.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Kuwait Foundation for the Advancement of Sciences (KFAS), "Towards Team-Based Cancer Care in Kuwait: Assessing the Continuum of Cancer Care" (CN20-

53MM-01). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

MA: Data curation; Formal analysis; Methodology; Visualization; Writing, original draft. AAt: Investigation (data acquisition and resources); Clinical interpretation; Writing, review and editing. DA: Methodology; Formal analysis (interpretation); Writing, review and editing. JT: Project administration; Writing, review and editing. WAR: Clinical interpretation; Supervision (clinical); Funding acquisition; Writing, review and editing. AAl: Conceptualization; Funding acquisition; Supervision; Methodology; Validation; Writing, review and editing. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the Information Technology Department at the Kuwait Cancer Control Center for their support in extracting and preparing the electronic health record data used in this study.

References

[1] GBD 2023 Cancer Collaborators. The global, regional, and national burden of cancer, 1990–2023, with forecasts to 2050: a systematic analysis for the Global Burden of Disease Study 2023. Lancet. 2025;406(10512):1565-86.

[2] Bizuayehu HM, Ahmed KY, Kibret GD, Dadi AF, Belachew SA, Bagade T, et al. Global disparities of cancer and its projected burden in 2050. JAMA Netw Open. 2024;7(11):e2443198.

[3] Craddock Lee SJ, Clark MA, Cox JV, Needles BM, Seigel C, Balasubramanian BA. Achieving coordinated care for patients with complex cases of cancer: a multiteam system approach. J Oncol Pract. 2016;12(11):1029-38.

[4] Pituskin E. Cancer as a new chronic disease: oncology nursing in the 21st century. Can Oncol Nurs J. 2022;32(1):87-92.

[5] Pizzoli SFM, Renzi C, Arnaboldi P, Russell-Edu W, Pravettoni G. From life-threatening to chronic disease: is this the case of cancers? A systematic review. Cogent Psychol. 2019;6(1):1-17.

[6] Taplin SH, Weaver S, Salas E, Chollette V, Edwards HM, Bruinooge SS, et al. Reviewing cancer care team effectiveness. J Oncol Pract. 2015;11(3):239-46.

[7] Taplin SH, Weaver S, Chollette V, Marks LB, Jacobs A, Schiff G, et al. Teams and teamwork during a cancer diagnosis: interdependency within and between teams. J Oncol Pract. 2015;11(3):231-8.

[8] Van Houdt S, Heyrman J, Vanhaecht K, Sermeus W, De Lepeleire J. An in-depth analysis of theoretical frameworks for the study of care coordination. Int J Integr Care. 2013;13(2):e024.

[9] McDonald KM, Sundaram V, Bravata DM, Lewis R, Lin N, Kraft S, et al. Closing the quality gap: a critical analysis of quality improvement strategies. Volume 7: care coordination. Rockville, MD: Agency for Healthcare Research and Quality; 2007. Technical Review No. 9. AHRQ Publication No. 04(07)-0051-7.

[10] Verhoeven DC, Chollette V, Lazzara EH, Shuffler ML, Osarogiagbon RU, Weaver SJ. The anatomy and physiology of teaming in cancer care delivery: a conceptual framework. J Natl Cancer Inst. 2021;113(4):360-70.

[11] Dai T, Tayur S. OM forum: healthcare operations management: a snapshot of emerging research. Manuf Serv Oper Manag. 2020;22(5):869-87.

[12] Nicolaisen A, Lauridsen GB, Haastrup P, Hansen DG, Jarbøl DE. Healthcare practices that increase the quality of care in cancer trajectories from a general practice perspective: a scoping review. Scand J Prim Health Care. 2022;40(1):11-28.

[13] Baker K, Dunwoodie E, Jones RG, Newsham A, Johnson O, Price CP, et al. Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. Int J Med Inform. 2017;103:32-41.

[14] Santos A, Leal GCL, Balancieri R. Process mining in healthcare: a tertiary study. BMC Med Inform Decis Mak. 2025;25(1):306.

[15] State of Kuwait. New Kuwait: Kuwait Vision 2035. Kuwait: General Secretariat of the Supreme Council for Planning and Development; 2016. https://www.newkuwait.gov.kw/home.aspx. Accessed 19 August 2026.

[16] Kuwait Cancer Control Center. KCCC strategic plan 2013–2018. Kuwait: Ministry of Health; 2013.

[17] Alhuwail D. Understanding health information management practices in public hospitals in Kuwait. Health Inf Manag. 2020;49(2–3):127-36.

[18] Alnashmi M, Salman A, AlHumaidi H, Yunis M, Al-Enezi N. Exploring the health information management system of Kuwait: lessons and opportunities. Appl Syst Innov. 2022;5(1):25.

[19] Hamadeh RR, Borgan SM, Sibai AM. Cancer research in the Arab world: a review of publications from seven countries between 2000–2013. Sultan Qaboos Univ Med J. 2017;17(2):e147-54.

[20] Zhang Y, Simoff MJ, Ost D, Wagner OJ, Lavin J, Nauman B, et al. Understanding the patient journey to diagnosis of lung cancer. BMC Cancer. 2021;21(1):402.

[21] Dusetzina SB, Tyree S, Meyer AM, Meyer A, Green L, Carpenter WR. Linking data for health services research: a framework and instructional guide. Rockville, MD: Agency for Healthcare Research and Quality; 2014. AHRQ Publication No. 14-EHC033-EF. https://www.ncbi.nlm.nih.gov/books/NBK253313/. Accessed 19 August 2026.

[22] 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.

[23] Crato N, Paruolo P, editors. Data-driven policy impact evaluation: how access to microdata is transforming policy design. Cham: Springer; 2019.

[24] Suriadi S, Andrews R, ter Hofstede AHM, Wynn MT. Event log imperfection patterns for process mining: towards a systematic approach to cleaning event logs. Inf Syst. 2017;64:132-50.

[25] van der Aalst WMP. Process mining: data science in action. 2nd ed. Berlin, Heidelberg: Springer; 2016.

[26] Fluxicon Process Laboratories. Disco, version 4.0.8 [computer software]; 2024. https://fluxicon.com/disco/. Accessed 19 August 2026. Eindhoven, the Netherlands.

[27] Janssenswillen G, Depaire B, Swennen M, Jans M, Vanhoof K. bupaR: enabling reproducible business process analysis. Knowl Based Syst. 2019;163:927-30.

[28] R Core Team. R: a language and environment for statistical computing, version 4.1.2; 2021. https://www.R-project.org/. Accessed 19 August 2026. R Foundation for Statistical Computing, Vienna, Austria.

[29] Günther CW, van der Aalst WMP. Fuzzy mining: adaptive process simplification based on multi-perspective metrics. In: Alonso G, Dadam P, Rosemann M, editors. Business Process Management (BPM 2007). vol. 4714 of Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2007. p. 328-43.

[30] Erdogan TG, Tarhan A. A goal-driven evaluation method based on process mining for healthcare processes. Appl Sci. 2018;8(6):894.

[31] Finch A, Cooper S, Raine R, Taylor RM, Gibson F. The development of ambulatory cancer care in the UK: a scoping review of the literature. Eur J Cancer Care. 2023;2023(1):4589362.