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.

Process mining of breast cancer care pathways from routine electronic medical records: coordination patterns, bottlenecks, and hospital-level variation

Ting Song
University of Wollongong
Su C. Ang
Macquarie University
Kimberley J. Davis
Illawarra Shoalhaven Local Health District
Chantal Campbell
University of Wollongong
Rebekah Costelloe
Illawarra Shoalhaven Cancer and Haematology Network, Illawarra Shoalhaven Local Health District
Ping Yu
University of Wollongong
Steven J. Craig
steven@drstevencraig.com.au

University of Wollongong

Research Article

Keywords: breast cancer, process mining, care pathways, electronic health records, care coordination, multidisciplinary care, clinical workflow

Posted Date: June 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9549473/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Breast cancer care involves multiple services over several months, and care pathways in routine practice vary widely between patients. Most previous research has examined treatment components separately, providing limited insight into how clinical events are sequenced and coordinated across services. Linked electronic medical record (EMR) data and process mining offer a way to reconstruct patient-level care pathways from data already captured in routine clinical systems. This study aimed to reconstruct breast cancer care pathways, identify coordination patterns and bottlenecks, and examine whether pathway organisation was associated with measured clinical characteristics or treating hospital.

Methods

A retrospective cohort study was conducted using linked EMR data from 688 patients with invasive breast cancer who underwent breast cancer surgery within the Illawarra Shoalhaven Local Health District, Australia, between 2010 and 2014. Pathways from diagnosis to treatment initiation were reconstructed from three routine clinical information systems using process mining. Coordination patterns were classified as MDT-led or Referral-led based on the relative timing of multidisciplinary team (MDT) discussion and oncology referral. Waiting times between consecutive care events were quantified, and multivariable logistic regression was used to identify predictors of coordination pattern and pre-treatment oncology consultation.

Results

The event log contained 4,184 timestamped events and 145 distinct pathway variants; 57.2% of variants occurred in only one patient. Despite this heterogeneity, most patients followed one of two coordination patterns: MDT-led, where MDT discussion occurred before or on the same date as oncology referral, or Referral-led, where oncology referral occurred before MDT discussion. Treating hospital was the only included model variable significantly associated with coordination pattern after adjustment (Shoalhaven vs Wollongong: OR 2.54, 95% CI 1.56–4.05). Molecular subtype, stage, grade, and age were not significantly associated. The longest observed intervals included surgery to oncology consultation and surgery to radiotherapy, with several key intervals longer at the regional hospital.

Conclusions

Linked EMR data and process mining can be used to reconstruct breast cancer care pathways, identify recurring coordination patterns, and quantify timing variation across multidisciplinary services. In this cohort, pathway organisation was more strongly associated with treating hospital than with measured

clinical characteristics, suggesting that routine clinical data can support pathway-based monitoring of coordination and timing variation in cancer care.

Background

Breast cancer is the most commonly diagnosed malignancy in women worldwide and remains a leading cause of cancer-related mortality, with an estimated 2.3 million new cases and 670,000 deaths globally in 2022 [1]. In Australia, where incidence rates are among the highest globally, breast cancer accounts for approximately 28% of all new cancer diagnoses, with an average of 58 women diagnosed each day [2]. Advances in surgery, systemic therapy, and radiotherapy have substantially improved survival. As outcomes improve, attention has increasingly shifted not only to treatment effectiveness, but also to how cancer care is organised, coordinated, and delivered across health systems [2, 3].

The digitisation of clinical workflows has made it increasingly feasible to study care delivery directly from data generated during routine practice [4–6]. Linked electronic medical record (EMR) data from surgical, oncology, and administrative systems contain time-stamped records of clinical activity that were previously accessible only through time-consuming manual review [7]. These records make it possible to examine not only what care was delivered, but how it was sequenced and coordinated across services. This is particularly relevant to cancer care, where treatment extends over months, involves multiple specialties, and varies substantially between patients and service settings [8].

Breast cancer care is inherently multidisciplinary. Patients typically move through diagnostic assessment, surgery, systemic therapy, radiotherapy, and multidisciplinary team (MDT) review, often across different services and clinicians [8–10]. In Australia, the Optimal Care Pathway for breast cancer identifies multidisciplinary care, coordination, communication, and supportive care as core components of high-quality care [11]. Although clinical guidelines specify which treatments are appropriate for a given tumour profile, they do not prescribe a single operational sequence through which care must be delivered [12]. In routine practice, variation may therefore arise not only from tumour biology and patient preference, but also from local workflow, referral arrangements, and the way care is documented in EMRs.

However, much of the existing literature on breast cancer services has focused on discrete care components, such as receipt of guideline-recommended therapy, treatment utilisation, or waiting times between selected milestones [10, 13]. Those studies have generated important evidence on quality and access, but they provide limited insight into how clinical events are sequenced and coordinated across the treatment trajectory, particularly in oncology settings where multiple services interact over time [14–17].

A key unanswered question is how real-world breast cancer care is organised when examined as a patient-level sequence of time-stamped events. It is unclear whether most patients follow a small number of recurring pathways or whether routine care is characterised by many distinct variants. It is also unclear whether this variation mainly reflects measured clinical characteristics or the way services

are organised. Answering these questions is important for using routine clinical data to monitor care coordination, identify avoidable delays, and inform service improvement, particularly in regional settings where access constraints and resource differences may amplify variation.

Process mining offers a practical way to examine these questions. Using time-stamped data already captured in health information systems, it reconstructs patient-level pathways and reveals how care is actually organised in practice [16, 18, 19]. This supports the identification of recurring pathway structures, variation in coordination, and bottlenecks in care delivery [16, 20]. Although process mining has been widely applied in healthcare, its use in complex longitudinal oncology settings remains limited, and breast cancer applications based on routinely collected EMR data are still emerging [15, 16, 21].

To address this gap, we used linked EMR data from an integrated public health district in regional Australia to reconstruct breast cancer care pathways during the first treatment episode. Specifically, we aimed to (1) characterise pathway structure and heterogeneity, (2) identify dominant coordination patterns and their determinants, (3) locate transition-specific bottlenecks within reconstructed pathways and examine their relation to coordination pattern and hospital context, and (4) assess whether measured clinical characteristics or hospital context were more strongly associated with pathway organisation.

Methods

Ethics Approval

This study was approved by the Joint Human Research Ethics Committee of the University of Wollongong and the Illawarra Shoalhaven Local Health District (ISLHD) (approval No. 2021/ETH00525). All data were deidentified before analysis, and no direct patient contact or intervention occurred. Data were stored and managed in accordance with the Australian National Statement on Ethical Conduct in Human Research.

Study Design and Setting

A retrospective observational cohort study was conducted within ISLHD, a publicly funded health network on the south coast of New South Wales, Australia. The district includes two acute hospitals: Wollongong Hospital (metropolitan) and Shoalhaven District Memorial Hospital (regional), operating under shared clinical and administrative governance. Breast cancer services across the district include diagnostic imaging, surgery, medical oncology, radiotherapy, and MDT review, supported by shared clinical information systems. This structure enabled the within-system comparison of pathway organisation across the two sites.

Data Source

Patient-level data were derived from three routinely used clinical information systems within ISLHD: SurgiNet (surgical and perioperative records), MOSAIQ (medical oncology and radiotherapy encounters), and the NSW Cancer Registry (verified diagnosis dates). All three systems were in routine clinical use during the study period, and no data were collected specifically for this study.

Eligible patients were female adults who underwent breast cancer surgery within ISLHD between January 2010 and December 2014 and had a primary diagnosis of invasive breast cancer (ICD-10 C50.0-C50.9) [22]. Patients were excluded if they had ductal carcinoma in situ only, a benign or unrelated diagnosis, recurrent disease diagnosed before 2010, prophylactic breast surgery without confirmed cancer, a secondary malignant breast tumour, solitary lymph node surgery without breast surgery, phyllodes tumour, neoadjuvant endocrine therapy, or initial surgery performed outside ISLHD.

The final analytical cohort comprised 688 patients. Related analyses of this cohort have been reported previously [23–25], and this study extends these works to reconstruct and analyse individual care pathways using process mining.

Data Preparation and Variable Standardisation

Several variables were standardised before pathway reconstruction to improve comparability across source systems. Molecular receptor status for oestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) was standardised across source systems into positive, negative, and unknown categories. Equivocal HER2 results without fluorescence in situ hybridisation confirmation were classified as unknown.

Molecular subtypes were then defined as luminal (ER-positive or PR-positive, HER2-negative), HER2-positive (HER2-positive regardless of hormone receptor status), triple-negative breast cancer (TNBC; ER-negative, PR-negative, and HER2-negative), and unknown. Luminal A and luminal B were not separated because Ki-67 data were unavailable, and hormone receptor status alone does not reliably distinguish between them [26].

Histological grade was standardised from mixed numeric and Roman numeral coding into ordinal categories 1, 2, and 3. Cases coded as metaplastic carcinoma (9) were retained as a separate category because histological grading is not applicable to this tumour type [27]. TNM stage was collapsed into four categories (I-IV) by merging substages.

Hormone therapy start dates were unavailable or incomplete for 98 patients with confirmed prescriptions. These patients were retained in the cohort, but their hormone therapy initiation events were excluded before pathway construction because the event dates could not be validly sequenced.

Date anomalies arising from earlier spreadsheet conversion errors were reviewed during data cleaning and corrected against source records where possible. Any unresolved dates were coded as missing before analysis.

Event Log Construction

Treatment pathways were reconstructed from an event log in which each patient was treated as a single case and each clinical activity as a dated event.

Based on clinical relevance and timestamp interpretability, eight event types were included: diagnosis (histological confirmation date), surgery (date of first breast cancer surgery), MDT discussion (date of MDT meeting), oncology referral (date of referral to medical or radiation oncology), oncology consultation (date of first oncology specialist consultation), chemotherapy initiation, radiotherapy initiation, and hormone therapy initiation. Re-excision for incomplete margin clearance and revision surgery were not counted as separate surgery events. Each event consisted of a case identifier, activity type, and event date. When multiple records of the same activity occurred on the same date for the same patient, they were collapsed into a single event to avoid inflating pathway complexity. Event dates were standardised to a consistent format before analysis.

Events were selected to preserve a clinically interpretable sequence. We included only activities that represented discrete treatment or coordination milestones and whose dates could be interpreted consistently across source systems. Activities that primarily reflected administrative or documentation processes, including admission, discharge, emergency department referral, and formal staging documentation, were excluded because their timestamps more often reflected data entry than the timing of the underlying clinical activity. For example, the recorded TNM staging date usually reflected formal documentation during oncology consultation, even when the pathological information used to assign stage had been available earlier. TNM stage was therefore retained as a baseline characteristic for descriptive and regression analyses, but not included as a sequence event.

Pathway analysis was restricted to the first treatment episode after diagnosis. Events related to recurrence, repeat surgery, or later treatment episodes were excluded to improve comparability across patients, because these episodes were more likely to reflect relapse, complications, or disease progression rather than variation in the initial treatment pathway.

Process Mining and Pathway Analysis

Process discovery was performed in Disco 4.2 (Fluxicon, Netherlands) [28]. To improve visual interpretability, frequency-based filtering was applied to the process map. A path threshold of 30% was selected after comparing maps generated at 20%, 30%, and 40%, as this level retained the dominant coordination structures while reducing infrequent branches that obscured the overall process map. This filtering affected only the graphical representation of the process map. Variant analysis, coordination pattern classification, and regression modelling were all performed on the complete unfiltered event log.

Variant analysis was conducted in Python (version 3.11) [29]. Each unique ordered sequence of events was treated as a distinct pathway variant. Variant frequencies were ranked, and cumulative patient coverage was calculated to describe pathway heterogeneity and long-tail structure.

Coordination Pattern Definition

Coordination pattern was defined by the relative timing of MDT discussion and oncology referral. Patients were classified as MDT-led if MDT discussion occurred before or on the same date as oncology referral, and as Referral-led if oncology referral occurred before MDT discussion. Same-day occurrences were classified as MDT-led in the primary analysis as an operational decision rule, reflecting the clinical plausibility that referral decisions may be initiated during MDT meetings [30]. This rule affected 73 patients (10.6%). Patients with only one of the two events recorded were classified as MDT-only or referral-only, and patients with neither event recorded were classified as having no recorded coordination.

This definition was based on two considerations. First, MDT discussion and oncology referral were treated as the principal coordination events for this analysis. MDT review is the point at which treatment recommendations are formulated, whereas referral marks entry into specialist systemic or radiation treatment planning. Their temporal order therefore provided a practical basis for distinguishing between two coordination patterns. Second, these were the two most consistently recorded coordination-related events in the dataset (98.7% and 82.1% of patients, respectively). Other potential markers of coordination, such as specialist consensus notes or nurse navigator activity, were not recorded consistently enough to support a cohort-wide operational classification.

Bottleneck Analysis

Bottleneck analysis was used to examine waiting times between consecutive care events. For each transition, we reported the median interval and IQR. For selected key intervals highlighted in the results, we additionally reported the proportion of patients waiting more than 30 or 60 days [31, 32]. These cut-points were used as descriptive markers to aid interpretation and were not intended to represent breast cancer-specific clinical thresholds in the Australian setting. Intervals were calculated only for patients in whom the later event occurred on or after the earlier event, so that reported waiting times reflected forward temporal flow.

Statistical Analysis

Descriptive statistics were used to summarise patient characteristics, event frequencies, and transition times. Continuous variables are reported as medians with IQRs, and categorical variables as counts and percentages. Comparisons of median transition times between two groups were performed using the Mann-Whitney U test, and comparisons across three or more groups using the Kruskal-Wallis test. Categorical differences were assessed using the chi-square test. Ninety-five percent confidence intervals for median transition times were estimated by bootstrap resampling with 2000 iterations. A two-sided P value of less than .05 was considered statistically significant [33].

Two multivariable logistic regression models were fitted [34]. Model A examined factors associated with pre-treatment oncology consultation, defined as an oncology consultation recorded on or before the first treatment date. Model B examined factors associated with MDT-led versus Referral-led coordination and was restricted to patients classified into either of these two groups. Both models included molecular subtype (reference: luminal), TNM stage (reference: stage I; stages III and IV were pooled because of small cell counts), histological grade (reference: grade 1), age group (reference: <65 years), and treating hospital (reference: Wollongong Hospital). Results are reported as odds ratios with 95% confidence intervals.

Model fit was assessed using the Hosmer-Lemeshow goodness-of-fit test and McFadden pseudo-R^2 [35]. Multicollinearity was assessed using variance inflation factors [36]. Because formal validation of timestamps against external workflow records was not available, robustness was assessed through predefined event selection rules and sensitivity analyses for the 73 patients with same-day MDT discussion and oncology referral [37]. Four alternative specifications were examined: same-day classified as MDT-led (primary analysis), same-day classified as Referral-led, same-day excluded, and a broader comparison of MDT-dominated patterns (MDT-led, same-day, and MDT-only) versus referral-dominated patterns (Referral-led and referral-only). In addition, a sensitivity analysis was performed with hormone therapy events removed from the event log to assess the effect of excluding hormone therapy initiation events for patients with missing hormone therapy dates. All analyses were performed in Python (version 3.11) using pandas, numpy, scipy, and matplotlib [29].

Results

The event log comprised 688 cases and 4,184 timestamped clinical events, with a mean of 6.1 events per patient. The median age at surgery was 65 years (IQR 54–74). Of the 688 patients, 153 (22.2%) were treated at Shoalhaven Hospital. Molecular subtype was classified as luminal in 530 patients (77.0%), HER2-positive in 79 (11.5%), TNBC in 65 (9.4%), and unknown in 14 (2.0%).

Overall Pathway Structure

The process map at a 30% path threshold showed the dominant pathway structure in the cohort (Fig. 1; see Appendix 1 for the unfiltered process map). Diagnosis and surgery were recorded for all 688 patients (100.0%). MDT discussion was recorded for 679 patients (98.7%), oncology consultation for 611 (88.8%), and oncology referral for 565 (82.1%). Treatment events included radiotherapy in 400 patients (58.1%), hormone therapy in 310 (45.1%), and chemotherapy in 243 (35.3%).

Eight treatment combinations were observed. The most frequent were surgery only (151/688, 21.9%), surgery with radiotherapy and hormone therapy (124/688, 18.0%), and surgery with radiotherapy alone (99/688, 14.4%).

Pathway Variants and Heterogeneity

The event log contained 145 unique pathway variants across 688 patients. Of these, 83 variants (57.2%) occurred with only one patient in each (Fig. 2A). The most frequent variant accounted for 8.9% of patients, and the 10 most frequent variants covered 45.5% of the cohort (Table 1). Pathway length ranged from 3 to 8 events, with a median of 6 (IQR 5–7) (Fig. 2B).

Table 1

Top 10 most frequent pathway variants in the reconstructed breast cancer event log (n = 688).

Rank	Patients, n (%)	Cumulative coverage (%)	Coordination pattern	Pathway sequence	No. of events
1	61 (8.9)	8.9	MDT-only	Diagnosis → Surgery → MDT discussion	3
2	52 (7.6)	16.4	MDT-led	Diagnosis → Surgery → MDT discussion → Oncology referral → Oncology consultation → Radiotherapy	6
3	41 (6.0)	22.4	MDT-led	Diagnosis → Surgery → MDT discussion → Oncology referral → Oncology consultation	5
4	32 (4.7)	27.0	Referral-led	Diagnosis → Surgery → Oncology referral → MDT discussion → Oncology consultation → Radiotherapy	6
5	28 (4.1)	31.1	Referral-led	Diagnosis → Surgery → Oncology referral → MDT discussion → Oncology consultation → Chemotherapy → Radiotherapy	7
6	26 (3.8)	34.9	MDT-led	Diagnosis → Surgery → MDT discussion → Oncology referral → Oncology consultation → Chemotherapy → Radiotherapy	7
7	19 (2.8)	37.6	MDT-led	Diagnosis → Surgery → MDT discussion → Oncology referral → Oncology consultation → Radiotherapy → Hormone therapy	7
8	19 (2.8)	40.4	MDT-led	Diagnosis → Surgery → MDT discussion → Oncology referral → Hormone therapy → Oncology consultation → Radiotherapy	7
9	19 (2.8)	43.2	MDT-led	Diagnosis → Surgery → MDT discussion → Oncology referral → Oncology consultation → Chemotherapy → Radiotherapy → Hormone therapy	8
10	16 (2.3)	45.5	Referral-led	Diagnosis → Surgery → Oncology referral → MDT	5

Note: Coordination pattern was defined according to the timing and/or presence of MDT discussion and oncology referral. Percentages were calculated using the full cohort denominator (n = 688) and rounded to one decimal place. Cumulative percentages were calculated from cumulative patient counts instead of summing rounded percentages.

Rank	Patients, n (%)	Cumulative coverage (%)	Coordination pattern	Pathway sequence	No. of events
				discussion → Oncology consultation	

Note: Coordination pattern was defined according to the timing and/or presence of MDT discussion and oncology referral. Percentages were calculated using the full cohort denominator (n = 688) and rounded to one decimal place. Cumulative percentages were calculated from cumulative patient counts instead of summing rounded percentages.

Coordination Patterns and Hospital-Level Differences

Two coordination patterns accounted for most patients in the cohort. In the MDT-led pattern, MDT discussion occurred before or on the same date as oncology referral in 333 patients (48.4%), including 73 (10.6%) with same-day events. In the Referral-led pattern, oncology referral preceded MDT discussion in 224 patients (32.6%). Oncology consultation followed the coordination events in both groups.

Among patients who received surgery only, 51.7% (78/151) had an MDT-only pattern without oncology referral, compared with 0.8%-3.4% across patients who received multimodal adjuvant therapy.

Measured patient characteristics, including age, molecular subtype, TNM stage, and histological grade, did not differ between the MDT-led and Referral-led groups (Table 2). The distribution of coordination patterns differed by hospital (Appendix 2; χ^2=30.1, P<.001), with MDT before referral and MDT-only patterns more common at Shoalhaven than at Wollongong.

Table 2
Patient characteristics and selected timing intervals by coordination pattern.

Variable	MDT-led (n = 333)	Referral-led (n = 224)	P value
Demographics			
Age at surgery, median (IQR)	64 (54–72)	65 (52–73)	.805
Age < 65, n (%)	173 (52.0)	110 (49.1)	.567
Age ≥ 65, n (%)	160 (48.0)	114 (50.9)	
Molecular subtype			
Luminal, n (%)	263 (79.0)	178 (79.5)	.735
HER2-positive, n (%)	34 (10.2)	19 (8.5)	
TNBC, n (%)	31 (9.3)	25 (11.2)	
Unknown, n (%)	5 (1.5)	2 (0.9)	
TNM stage			
Stage I, n (%)	138 (41.4)	108 (48.2)	.447
Stage II, n (%)	147 (44.1)	85 (37.9)	
Stage III, n (%)	45 (13.5)	29 (12.9)	
Stage IV, n (%)	3 (0.9)	2 (0.9)	
Histological grade			
Grade 1, n (%)	62 (18.6)	43 (19.2)	.960
Grade 2, n (%)	151 (45.3)	105 (46.9)	
Grade 3, n (%)	115 (34.5)	75 (33.5)	
Pathway timing			
Surgery to first treatment, median days (IQR)	56 (42–76)	49 (36–63)	< .001
Surgery to oncology consultation, median days (IQR)	36 (25–49)	28 (21–36)	< .001

Note: Percentages are column percentages within each coordination group. P values refer to overall comparisons within each variable block. Patients with missing or non-applicable histological grade were omitted from the histological grade rows and excluded from the grade comparison. For timing variables, P values compare MDT-led and Referral-led groups. Hospital distribution by coordination pattern is reported in Appendix 2.

In multivariable analysis, treating hospital was the only included model variable significantly associated with coordination pattern after adjustment (Shoalhaven vs Wollongong: OR 2.54, 95% CI 1.56–4.05; P<.001; Fig. 3, Model B). Molecular subtype, stage, grade, and age were not significantly associated with coordination pattern.

Model B showed adequate calibration (Hosmer-Lemeshow χ^2=8.1, df = 8, P=.426), although overall explanatory power was low (McFadden pseudo-R^2=0.027). The association with treating hospital remained stable across sensitivity analyses using alternative approaches to same-day MDT discussion and oncology referral (adjusted OR 2.54-3.00; all P<.001; Appendix 3). Excluding hormone therapy events in a separate sensitivity analysis did not materially change the coordination pattern distribution or the association with hospital (Appendix 4).

Temporal Analysis and Bottlenecks

The longest observed intervals occurred after surgery. The median interval from surgery to oncology consultation was 33 days (IQR 22–43), with 53.8% (322/598) of patients waiting more than 30 days. Among patients who received radiotherapy, the median interval from surgery to radiotherapy was 100 days (IQR 69–186), with 88.8% waiting more than 60 days (Table 3, Fig. 4). As this interval may include chemotherapy or other intervening treatment steps, it should not be interpreted as a direct measure of radiotherapy access delay alone.

Table 3
Waiting times between key care events, overall and by hospital or coordination pattern.

Interval / subgroup	n	Median (days)	IQR	95% CI	P value
Diagnosis to surgery					
Overall	688	17	14–27	16–20	—
Wollongong Hospital	535	16	13–23	15–16	< .001
Shoalhaven Hospital	153	27	19–34	22–28	
Surgery to oncology consultation					
Overall	598	33	22–43	30–35	—
Wollongong Hospital	468	32	25–43	29–35	.230
Shoalhaven Hospital	130	34	16–45	29–38	
MDT-only	41	31	24–40	27–36	< .001
MDT-led	331	36	25–49	35–39	
Referral-led	217	28	21–36	26–29	
Surgery to chemotherapy					
Overall	238	50	42–64	48–54	—
Wollongong Hospital	181	48	40–62	46–50	< .001
Shoalhaven Hospital	57	63	50–72	55–64	
Surgery to radiotherapy					
Overall	400	100	69–186	89–117	—
Wollongong Hospital	326	96	68–179	88–116	< .001
Shoalhaven Hospital	74	160	85–203	105–181	
Surgery to first treatment initiation					
Overall	526	55	40–70	52–56	—
MDT-only	42	59	41–79	47–66	< .001
MDT-led	281	56	42–76	54–62	
Referral-led	194	49	36–63	46–55	

Note: Intervals were calculated only for patients in whom the later event occurred on or after the earlier event. P values compare subgroups within each transition block: Wollongong vs Shoalhaven for hospital comparisons, and MDT-only, MDT-led, and Referral-led groups for coordination-pattern comparisons. The interval from surgery to radiotherapy should not be interpreted as a direct measure of radiotherapy waiting time because it may include chemotherapy or other intervening treatment.

Among patients who received chemotherapy followed by radiotherapy, the median interval from chemotherapy to radiotherapy was 145 days (IQR 122–193).

Inter-hospital differences were observed across several key transitions. Compared with Wollongong, patients treated at Shoalhaven waited longer from diagnosis to surgery (median 27 vs 16 days, P<.001), surgery to chemotherapy (63 vs 48 days, P<.001), and surgery to radiotherapy (160 vs 96 days, P<.001; Table 3).

Timing also differed by coordination pattern. The median interval from surgery to oncology consultation was shorter in the Referral-led group than in the MDT-led group (28 vs 36 days, P<.001), and the median interval from surgery to first treatment was also shorter (49 vs 56 days, P<.001; Table 2).

Among patients who received radiotherapy, the interval from surgery to radiotherapy differed by molecular subtype (median 86 days in luminal disease, 182 days in HER2-positive disease, and 207 days in TNBC; P<.001). This interval was pathway-dependent and could include chemotherapy before radiotherapy.

Exploratory Predictors of Pre-Treatment Oncology Consultation

The proportion of patients seen by oncology before first treatment differed by molecular subtype, ranging from 79.1% in luminal disease to 94.3% in triple-negative breast cancer (P=.019; Fig. 3, Model A). In multivariable analysis, grade 3 histology was associated with pre-treatment oncology consultation (OR 2.55, 95% CI 1.20–5.39; P=.015), whereas other variables were not statistically significant. Model A had low explanatory power (McFadden pseudo-R^2=0.049) and borderline calibration (Hosmer-Lemeshow χ^2=16.8, df = 8, P=.033).

Discussion

Principal Results

Although process mining is increasingly used in clinical research [15, 38, 39], its application to breast cancer care pathways reconstructed from routinely collected EMR data remains limited [21]. In this study, we used linked EMR data from an Australian public health network to reconstruct patient-level breast cancer pathways and examine their structure, variation, and timing within a single service system. Three main findings emerged. First, care pathways were highly heterogeneous at the patient level, with 145 distinct variants identified among 688 patients, and more than half occurred in only one patient. Second, despite this heterogeneity, most patients followed one of two dominant coordination patterns, defined by whether oncology referral occurred before or after MDT discussion. Third, treating hospital was the only model variable associated with coordination pattern after adjustment, whereas the measured clinical characteristics included in the model were not associated. This hospital association remained stable across sensitivity analyses, although unmeasured confounding cannot be excluded.

These findings add to the literature that has largely examined breast cancer services in terms of treatment components, guideline adherence, or waiting times between selected milestones [10, 40–42]. Those approaches provide important evidence on quality and access, but they do not capture how care is sequenced across multiple services over time. By reconstructing care as a longitudinal series of time-stamped events, this study shows how linked EMR data can be used to identify pathway heterogeneity, recurring coordination patterns, and transition-specific timing variation within a multidisciplinary cancer service [20, 21]. This study extends recent process-mining work in breast cancer care based on routine data by examining coordination patterns, hospital-level differences, and their relation to treatment timing within a linked EMR dataset [21]. It also extends earlier work using the same dataset [23]. These findings show that linked EMR data collected for clinical care can be used to reconstruct pathway structure, identify coordination patterns, and quantify waiting-time variation across a multidisciplinary cancer service.

Pathway Heterogeneity and Higher-Level Coordination Patterns

Although breast cancer is often regarded as a relatively standardised cancer type with treatment recommendations based on tumour biology [12], the organisation of care pathways is often much less standardised. Guidelines specify which treatments are appropriate, but they do not define a single operational sequence through which coordination, referral, and treatment must occur. This is consistent with the Australian optimal care pathway, which frames breast cancer care as a coordinated multidisciplinary process while recognising that not all patients follow every step linearly [11].

Substantial pathway heterogeneity was observed in this study. Nearly half of the cohort was concentrated within the top ten most frequent variants, indicating that pathway variation was not random but clustered within a limited number of recurring structures. Similar forms of structured heterogeneity have been reported in oncology and other complex clinical settings [17, 43].

Regularity was also evident at a broader level. Although 145 patient-level variants were identified, two coordination patterns accounted for most patients, and both converged on oncology consultation. This suggests that variation in routine care occurred within a limited number of recurring service-level coordination patterns [8, 44].

Coordination Pattern and Hospital Context

A key finding was the association between treating hospital and coordination pattern. Molecular subtype, stage, grade, and age were not significantly associated with whether patients followed an MDT-led or Referral-led sequence, whereas treating hospital remained significantly associated across the primary model and sensitivity analyses. This does not exclude a role for unmeasured clinical or contextual factors, but it suggests that the measured clinical characteristics included in this study did not account for the observed differences in pathway organisation.

One interpretation is that the observed difference in coordination pattern reflects differences in how MDT review and oncology referral were operationalised across the two hospitals. Although the available data do not identify the underlying workflow directly, the findings are consistent with local differences in the sequencing of these coordination steps in routine practice. This is consistent with the concept of unwarranted variation in health services research, in which differences in care delivery reflect service organisation as well as patient need [45]. In this setting, such variation may indicate areas for future workflow review and redesign. This is relevant in the Australian context, where the Optimal Care Pathway identifies multidisciplinary care and care coordination as core principles of breast cancer care and recommends MDT discussion for newly diagnosed patients before surgery or neoadjuvant chemotherapy [11].

Same-day MDT discussion and oncology referral required particular attention because these timestamps may reflect either genuine same-day decision making or same-day data entry. For individual event pairs, the available data cannot distinguish between these possibilities. At the cohort level, however, the association with hospital changed little across sensitivity analyses, suggesting that the observed hospital-level difference in coordination pattern was not materially affected by uncertainty in the ordering of same-day events. This highlights the importance of checking how pathway definitions behave under plausible timestamp ambiguity when using routine clinical data.

Time Intervals and Inter-Hospital Differences

The timing analyses add a second dimension to the pathway findings. The median interval from surgery to oncology consultation was 33 days, with more than half of the patients waiting longer than 30 days. The interval from surgery to radiotherapy was longer, with a median of 100 days, although this measure was pathway-dependent and could include intervening chemotherapy. Across several key transitions, patients treated at Shoalhaven had longer intervals than those treated at Wollongong. This pattern was observed from diagnosis to surgery and from surgery to chemotherapy and radiotherapy. These differences were not evenly distributed across hospitals and were consistent with previous evidence of regional inequalities in access to cancer care [23].

Considered alongside the coordination analyses, these findings suggest that hospital context was associated with both the ordering of care events and the duration of selected treatment transitions within the same service setting.

The shorter median time to first treatment in the Referral-led group (49 days) than in the MDT-led group (56 days) might appear to favour direct referral pathways on efficiency grounds. However, this difference does not permit an inference about care quality. MDT discussion may serve functions beyond pathway sequencing, including treatment planning, multidisciplinary input on borderline indications, and documentation of clinical reasoning. Whether the shorter interval observed in Referral-led pathways translates into equivalent or better outcomes will require linkage to recurrence and survival data.

Clinical Predictors of Pre-treatment Consultation

Among the measured clinical variables, grade 3 histology was the only variable independently associated with pre-treatment oncology consultation. TNBC showed a positive but non-significant association after adjustment. Given the low explanatory power and borderline calibration of Model A, this finding should be regarded as preliminary and hypothesis-generating. By contrast, the more robust findings in this study were the coordination patterns identified in Model B and the transition-based timing analyses, in which the association with hospital remained consistent across sensitivity analyses.

Interpretation of the Regression Results

These regression models were used to examine whether the selected clinical variables and treating hospital were associated with pathway organisation. The low pseudo-R^2 values indicate that much of the observed variation was not explained by the variables available in this dataset. The models therefore provide a partial account of pathway organisation. For Model A, the borderline Hosmer-Lemeshow result further supports cautious interpretation of the association with grade 3 histology.

Implications for EMR-Based Service Monitoring

A broader implication of this study is that pathway-based analyses can be conducted using data already captured in routine clinical systems, without prospective data collection. In this cohort, linked EMR data were sufficient to reconstruct care pathways, compare coordination patterns, and quantify timing variation across key transitions. These findings suggest that process mining could support not only retrospective research, but also service monitoring and local quality improvement. This is consistent with the intended role of the breast cancer Optimal Care Pathway as a health-system tool for identifying service gaps, improving coordination, and assessing adherence where possible [11]. Similar approaches may also be applicable to other tumour streams or clinical settings in which care is distributed across multiple teams and information systems.

Limitations

This study has limitations. First, the findings were derived from a single integrated health district, which may limit generalisability. However, the inclusion of both a metropolitan and a regional hospital within the same governance structure allowed direct within-system comparison of pathway organisation.

Second, the study period predates more recent changes in digital health infrastructure and cancer service delivery. Although contemporary workflows may differ, the dataset provides a historically complete linked record of the first treatment episode, which is useful for pathway reconstruction. Replication using more recent datasets and additional service settings is needed to assess the contemporary applicability of these findings.

Third, the analysis relied on routinely collected EMR data, so timestamps reflected recorded event dates and not the exact timing of clinical activity. The event log was therefore restricted to eight event types with comparatively interpretable and reliable timestamps. The 145 pathway variants should be read in

that context: they reflect structural heterogeneity within this event-definition framework, but do not capture all clinical workflow variation [46, 47].

Fourth, several clinically relevant variables, including comorbidity, performance status, and patient treatment preferences, were unavailable. These factors may have influenced both pathway organisation and treatment timing. Hormone therapy initiation dates were also incomplete for a subset of patients, so those events could not be fully incorporated into pathway reconstruction. Although sensitivity analyses excluding hormone therapy events did not materially change the main findings, residual bias related to missing event dates and unmeasured confounding cannot be excluded.

Finally, the association with hospital should be understood in the context of a comparison between two hospitals within a single health district and should not be interpreted as evidence of a generalisable institution-level effect across health systems. Some predictors were also simplified for modelling, including age-based dichotomisation and pooling of stage III and IV disease, which may have reduced model granularity. In addition, observational design does not support causal inference. Future studies linking pathway structure with recurrence and survival outcomes are needed to determine whether the differences observed here translate into meaningful differences in treatment effectiveness.

Conclusions

Breast cancer care pathways in this health system were highly heterogeneous at the patient level, but the ordering of key coordination events differed between hospitals. Among the variables included in the model, coordination pattern was more strongly associated with treating hospital than with the measured clinical characteristics of patients, and this association remained consistent across multivariable adjustment and sensitivity analyses. These findings are consistent with hospital-level differences in pathway organisation, although substantial unexplained variation remained.

By combining pathway reconstruction with coordination pattern classification, transition-based timing analysis, and multivariable regression, this study shows how linked EMR data can be used to examine how cancer care is organised in routine practice. The inter-hospital differences in treatment intervals identified here may represent targets for service review and improvement. Future research should link pathway structure with recurrence and survival outcomes to determine whether the coordination differences and timing variation identified here translate into differences in treatment effectiveness.

Abbreviations

AIHW
Australian Institute of Health and Welfare
CI
Confidence interval
EMR
Electronic medical record

ER
Oestrogen receptor
HER2
Human epidermal growth factor receptor 2
IQR
Interquartile range
ISLHD
Illawarra Shoalhaven Local Health District
MDT
Multidisciplinary team
OR
Odds ratio
PR
Progesterone receptor
TNBC
Triple-negative breast cancer
TNM
Tumour, node, metastasis

Declarations

Ethics approval and consent to participate

This study was approved by the Joint Human Research Ethics Committee of the University of Wollongong and the Illawarra Shoalhaven Local Health District (approval No. 2021/ETH00525). The requirement for individual informed consent was waived because the study used deidentified routinely collected clinical data and involved no direct patient contact or intervention.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are not publicly available due to ethics, privacy, and data governance restrictions relating to patient-level clinical records. Data may be available from the corresponding author on reasonable request and subject to approval by the relevant ethics committee and data custodian.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by a Cancer Institute NSW Innovations in Cancer Control Grant 2022 (2022/INN1259) and a 2023 University of Wollongong Rural Health Research Grant (M2079). The funders had no role in the study design, data collection, analysis, interpretation of data, manuscript preparation, or the decision to submit the manuscript for publication.

Authors' contributions

SJC contributed to study conceptualisation, methodology, funding acquisition, validation, supervision, clinical interpretation, and manuscript revision. TS contributed to study design, methodology, funding acquisition, data curation, formal analysis, and drafting and revision of the manuscript. SCA contributed to clinical interpretation and manuscript revision. KJD contributed to methodology, funding acquisition, investigation, data curation, formal analysis, supervision, and manuscript revision. CC and RC contributed to investigation, data curation, and manuscript revision. PY contributed to study conceptualisation, methodology, funding acquisition, supervision, and manuscript revision. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Kira Jones and Sue Bartley for their assistance with data extraction and cleaning.

Authors' information

Not applicable.

References

1. Kim J, et al. Global patterns and trends in breast cancer incidence and mortality across 185 countries. Nat Med. 2025;31(4):1154–62.
2. Australian Institute of Health and Welfare, Cancer data in Australia. 2024, AIHW: Retrieved from https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia
3. Wang J, Wu S-G. Breast cancer: an overview of current therapeutic strategies, challenge, and perspectives. Breast Cancer: Targets Therapy, 2023: pp. 721–30.

4. Sicotte C, et al. Benefits of improving processes in cancer care with a care pathway-based electronic medical record. Practical radiation Oncol. 2016;6(1):26–33.
5. Berger ML, et al. Opportunities and challenges in leveraging electronic health record data in oncology. Future Oncol. 2016;12(10):1261–74.
6. Khozin S, Blumenthal GM, Pazdur R. Real-world data for clinical evidence generation in oncology. JNCI: J Natl Cancer Inst. 2017;109(11):djx187.
7. Kurian AW, et al. Breast cancer treatment across health care systems: linking electronic medical records and state registry data to enable outcomes research. Cancer. 2014;120(1):103–11.
8. Katano A, et al. The critical role of multidisciplinary teams in health care: Perspectives from radiation oncology. J Clin Sci Res. 2025;14(4):242–7.
9. Cohen-Stavi CJ, et al. Mixed Methods Evaluation of Reasons Why Care Deviates From Clinical Guidelines Among Patients With Multimorbidity. Med Care Res Rev. 2022;79(1):102–13.
10. Barth JH, et al. Why are clinical practice guidelines not followed? Clin Chem Lab Med (CCLM). 2016;54(7):1133–9.
11. Cancer Council Victoria. Victoria. Department of Health. Optimal care pathway for people with breast cancer. Cancer Council Victoria; 2021.
12. Gradishar WJ, et al. Breast Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2022;20(6):691–722.
13. Bleicher RJ, et al. Time to surgery and breast cancer survival in the United States. JAMA Oncol. 2016;2(3):330–9.
14. Ma SJ, et al. Association of Timing of Adjuvant Therapy With Survival in Patients With Resected Stage I to II Pancreatic Cancer. JAMA Netw Open. 2019;2(8):e199126.
15. Aversano L, et al. Process mining applications in healthcare: a systematic literature review. PeerJ Comput Sci. 2025;11:e2613.
16. Guzzo A, Rullo A, Vocaturo E. Process mining applications in the healthcare domain: A comprehensive review. Wiley Interdisciplinary Reviews: Data Min Knowl Discovery. 2022;12(2):e1442.
17. Santos A, Leal GCL, Balancieri R. Process mining in healthcare: a tertiary study. BMC Med Inf Decis Mak. 2025;25(1):306.
18. Dagliati A, et al. A Process Mining Pipeline to Characterize COVID-19 Patients' Trajectories and Identify Relevant Temporal Phenotypes From EHR Data. Front Public Health. 2022;10:815674.
19. Shafei I, Karnon J, Crotty M. Process mining and customer journey mapping in healthcare: Enhancing patient-centred care in stroke rehabilitation. Digit health. 2024;10:20552076241249264.
20. De Roock E, Martin N. Process mining in healthcare–An updated perspective on the state of the art. J Biomed Inform. 2022;127:103995.
21. Ferré F, et al. Enhancing quality and decision-making for care pathways: An application of process mining in cancer care. PLoS ONE. 2026;21(1):e0339788.

22. Abraha I, et al. Accuracy of administrative databases in detecting primary breast cancer diagnoses: a systematic review. BMJ Open. 2018;8(7):e019264.
23. Davis KJ, et al. Causes of unwarranted variation and disparity in breast cancer management in regional and rural area. Breast J. 2024;2024(1):9354395.
24. Song T et al. The importance of timely multidisciplinary teams involvement in breast cancer care. in ANZ J Surg. 2024. p. 246.
25. Campbell C, et al. Causes of unwarranted variation in breast cancer management in regional and rural areas. ANZ J Surg. 2024;94(S1):5.
26. Nielsen TO et al. *Assessment of Ki67 in Breast Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group.* JNCI: Journal of the National Cancer Institute, 2021. 113(7): pp. 808–819.
27. Altaf FJ, et al. Metaplastic carcinoma of the breast: an immunohistochemical study. Diagn Pathol. 2014;9(1):139.
28. Günther CW, Rozinat A. *Disco: Discover your processes*. in *Demonstration Track of the 10th International Conference on Business Process Management, BPM Demos 2012*. 2012. CEUR-WS. org.
29. Wang H-I, Manolas C, Xanthidis D. *Statistical analysis with Python*, in *Handbook of computer programming with Python*. 2022, Chapman and Hall/CRC. pp. 373–408.
30. Wihl J et al. Contributions to multidisciplinary team meetings in cancer care: predictors of complete case information and comprehensive case discussions. J multidisciplinary Healthc, 2021: pp. 2445–52.
31. Neal R, et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. Br J Cancer. 2015;112(1):S92–107.
32. Hanna TP et al. Mortality due to cancer treatment delay: systematic review and meta-analysis. BMJ, 2020. 371.
33. McCrum-Gardner E. Which is the correct statistical test to use? Br J Oral Maxillofac Surg. 2008;46(1):38–41.
34. Qu Y. Evaluation of a surrogate marker: validity and efficiency. Stat Med. 2013;32(12):1995–2000.
35. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006;25(1):127–41.
36. O'brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Volume 41. Quality & Quantity; 2007. pp. 673–90. 5.
37. Thabane L, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC Med Res Methodol. 2013;13(1):92.
38. Qian S, Song T, Yu P. *A Study Design to Model Clinical Management Process for the Health and Wellbeing of Patients with Alcohol Use Disorders*, in *MedInfo*. 2019. pp. 1753–1754.
39. Hitige NPR, et al. Appendicectomy pathway: Insights from electronic medical records of a local health district in Australia. Surgery. 2024;176(4):1001–7.

40. Vigneulles BSd., editor. *Systemic Modeling of Care Pathways by Identifying Invariant Determinants: A Qualitative Study Protocol.* International Journal of Qualitative Methods, 2024. 23: p. 16094069241251543.
41. Ewart E, et al. Assurance of timely access to breast cancer diagnosis and treatment by a regional breast health clinic serving both urban and rural-remote communities. Curr Oncol. 2023;30(1):1232–42.
42. Krause A, et al. Guideline adherence and implementation of tumor board therapy recommendations for patients with gastrointestinal cancer. J Cancer Res Clin Oncol. 2023;149(3):1231–40.
43. Munoz-Gama J, et al. Process mining for healthcare: Characteristics and challenges. J Biomed Inform. 2022;127:103994.
44. Grajek M, et al. The role and competencies of an oncology care coordinator: an essential link in patient-centered cancer care. Rep Practical Oncol Radiotherapy. 2025;30(5):619–27.
45. Wennberg JE. Unwarranted variations in healthcare delivery: implications for academic medical centres. BMJ. 2002;325(7370):961–4.
46. Mohamed Y, et al. Electronic health record data quality variability across a multistate clinical research network. J Clin Transl Sci. 2023;7(1):e130.
47. Penev YP, et al. Electronic Health Record Data Quality and Performance Assessments: Scoping Review. JMIR Med Inf. 2024;12:e58130.

Figures

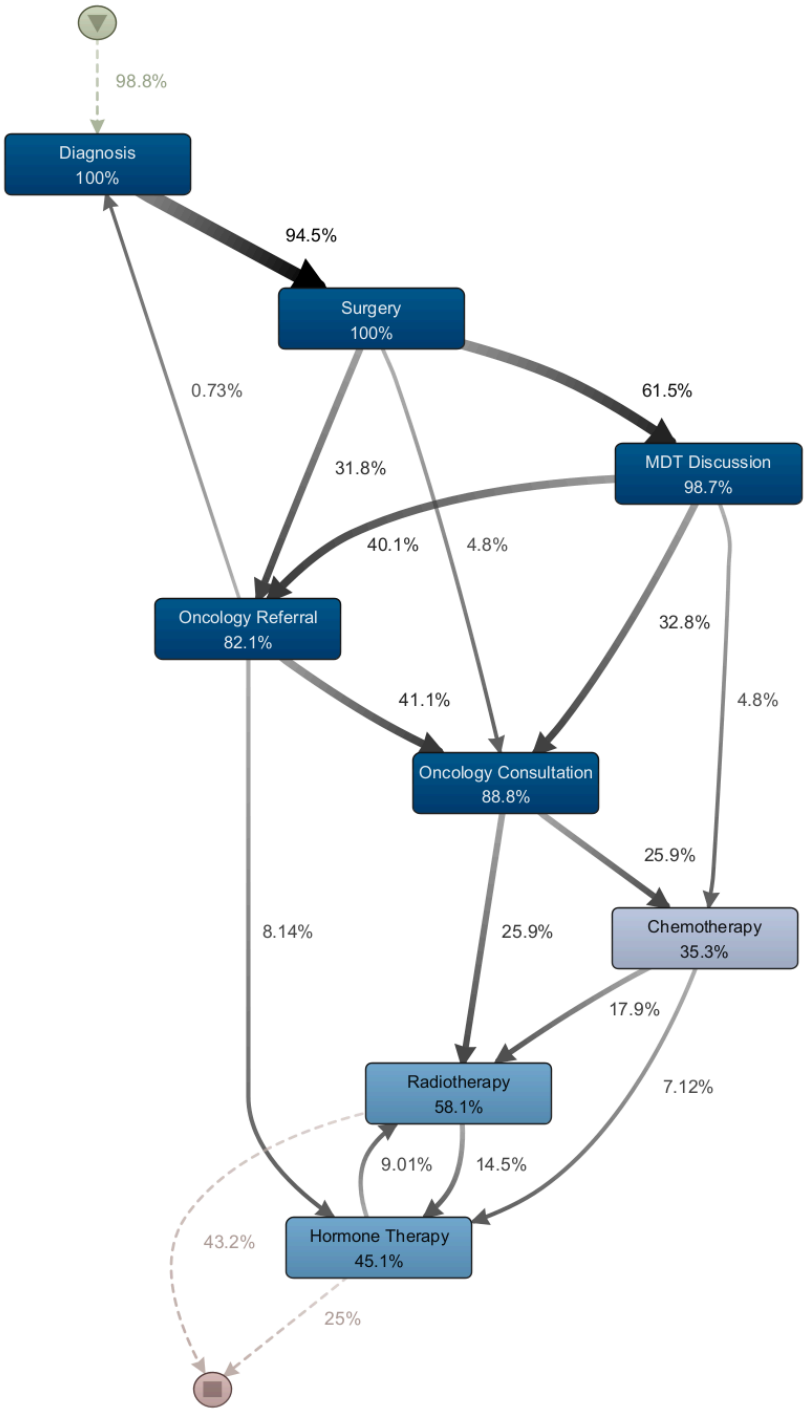


Figure 1

Process map of breast cancer care pathways at 30% path threshold

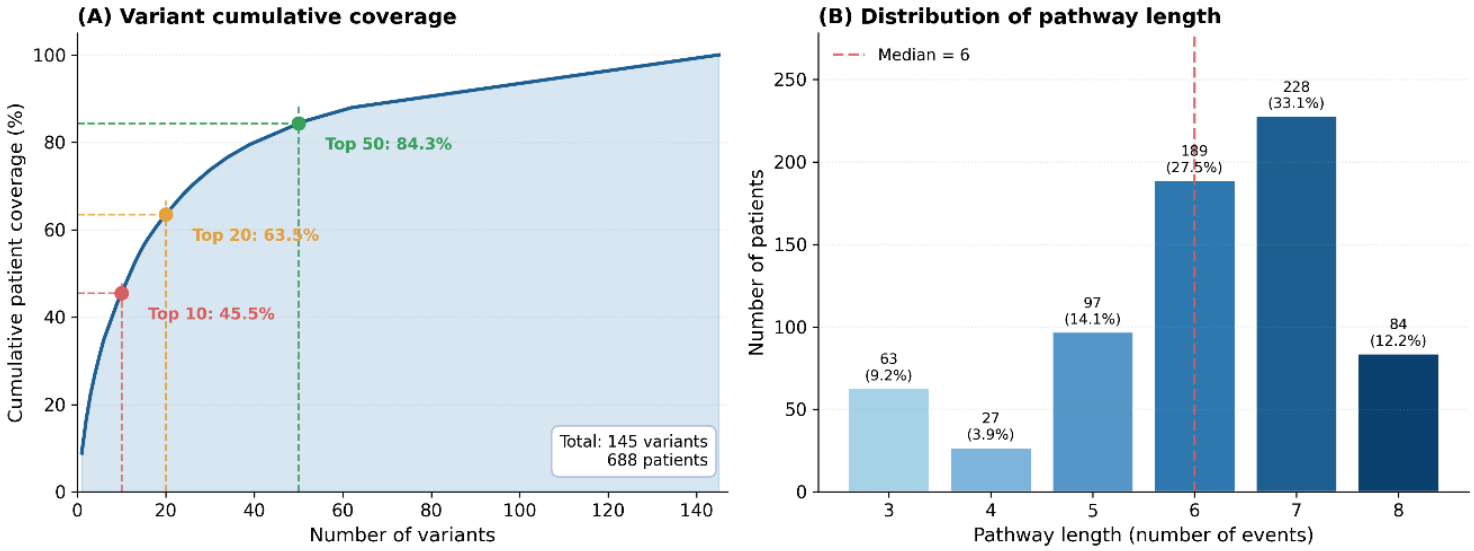


Figure 2

Pathway variant analysis

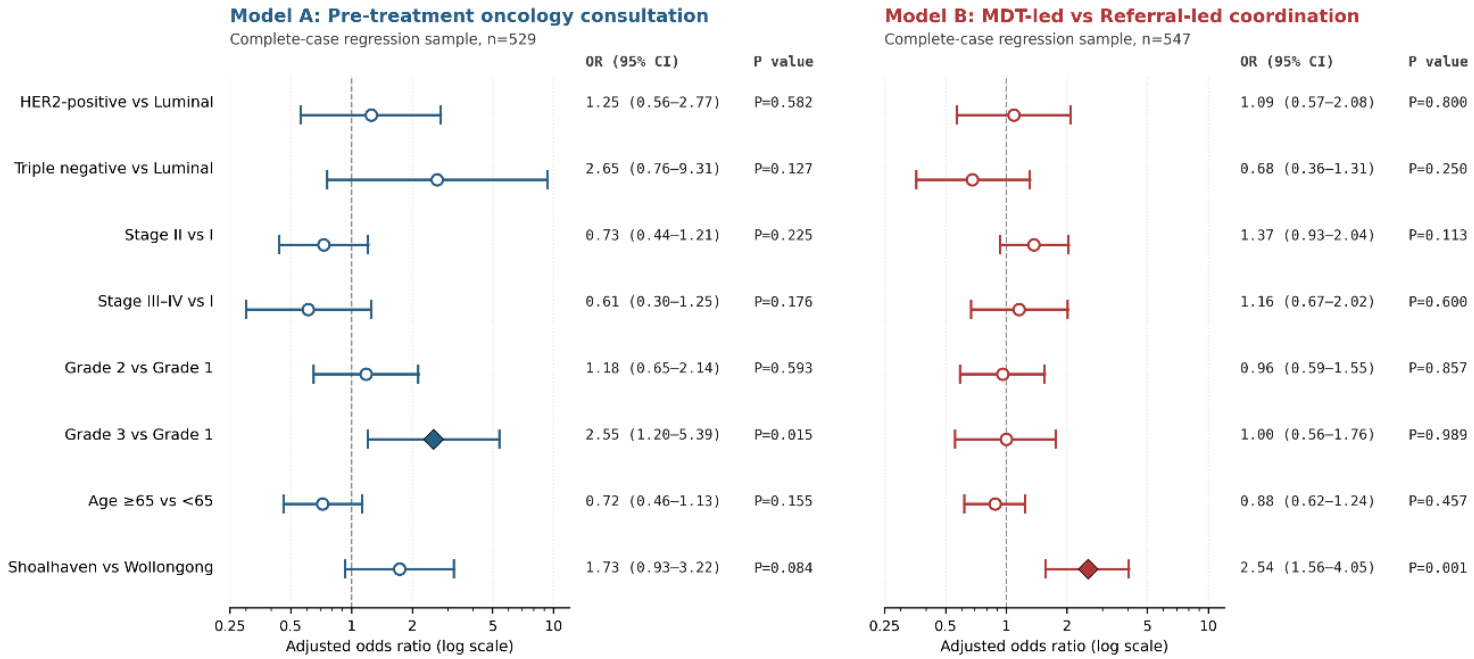


Figure 3

Adjusted odds ratios from multivariable logistic regression models for pre-treatment oncology consultation (Model A; complete-case regression sample, n=529) and MDT-led vs Referral-led coordination (Model B; complete-case regression sample, n=547).

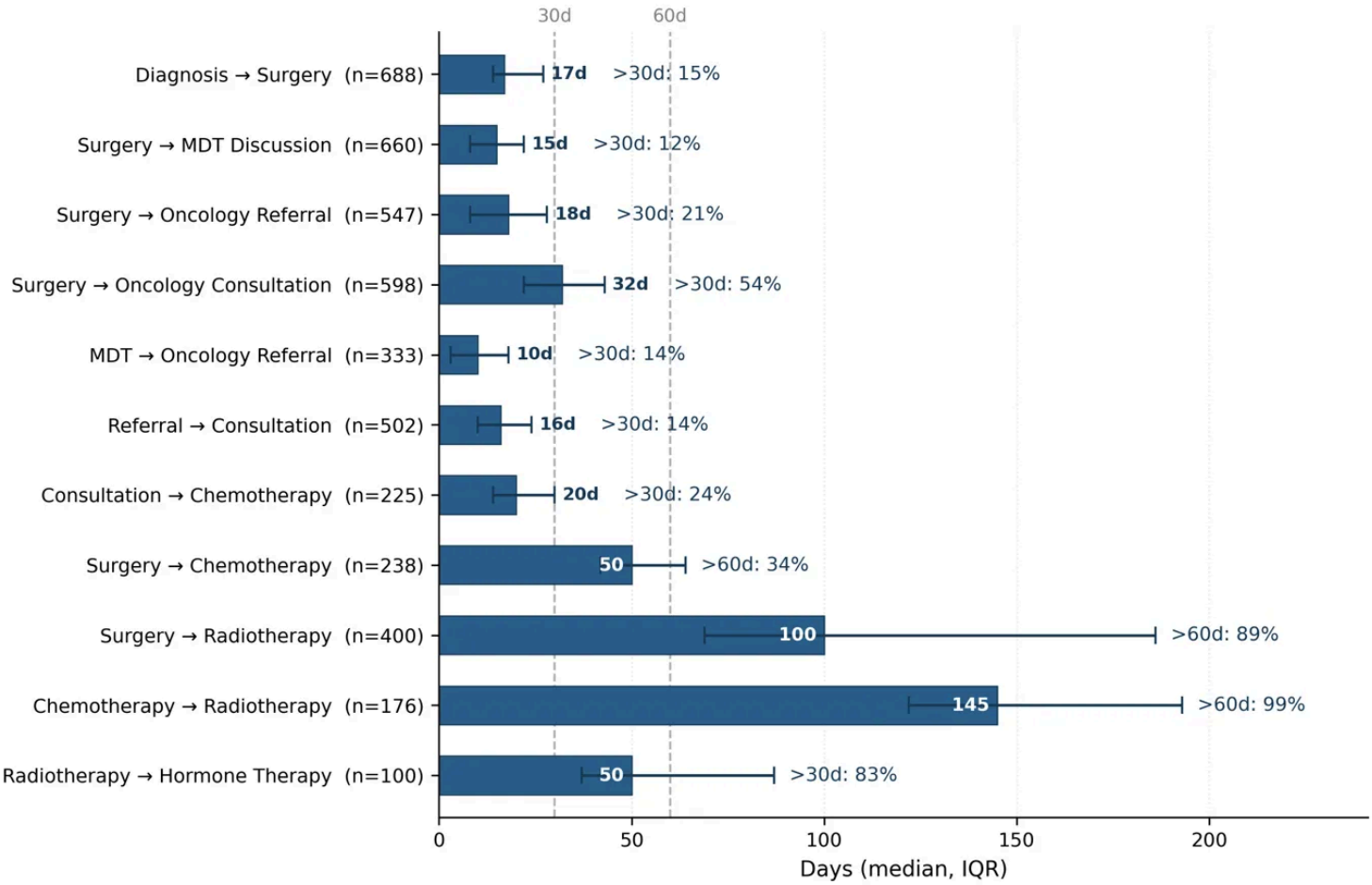


Figure 4

Bottleneck analysis of waiting times between consecutive care events.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx