






Estimating Treatment Effect of Adjuvant Chemotherapy in Elderly Patients With Stage III Colon Cancer Using Bayesian Networks

Melle Sieswerda, MD^{1,2} ; Ruby van Rossum, MSc¹; Inigo Bermejo, PhD²; Gijs Geleijnse, PhD¹ ; Katja Aben, PhD^{1,3} ; Felice van Erning, PhD^{1,4} ; Ignace de Hingh, MD, PhD^{1,4}; Valery Lemmens, PhD²; André Dekker, PhD² ; and Xander Verbeek, PhD¹

DOI <https://doi.org/10.1200/CCI.23.00080>

ABSTRACT

PURPOSE While adjuvant therapy with capecitabine and oxaliplatin (CAPOX) has been proven to be effective in stage III colon cancer, capecitabine monotherapy (CapMono) might be equally effective in elderly patients. Unfortunately, the elderly are under-represented in clinical trials and patients included may not be representative of the routine care population. Observational data might alleviate this problem but is sensitive to biases such as confounding by indication. Here, we build causal models using Bayesian Networks (BNs), identify confounders, and estimate the effect of adjuvant chemotherapy using survival analyses.

METHODS Patients 70 years and older were selected from the Netherlands Cancer Registry (N = 982). We developed several BNs using constraint-based, score-based, and hybrid algorithms while precluding noncausal relations. In addition, we created models using a limited set of recurrence and survival nodes. Potential confounders were identified through the resulting graphs. Several Cox models were fitted correcting for confounders and for propensity scores.

RESULTS When comparing adjuvant treatment with surgery only, pathological lymph node classification, physical status, and age were identified as potential confounders. Adjuvant treatment was significantly associated with survival in all Cox models, with hazard ratios between 0.39 and 0.45; CIs overlapped. BNs investigating CAPOX versus CapMono did not find any association between the treatment choice and survival and thus no confounders. Analyses using Cox models did not identify significant association either.

CONCLUSION We were able to successfully leverage BN structure learning algorithms in conjunction with clinical knowledge to create causal models. While confounders differed depending on the algorithm and included nodes, results were not contradictory. We found a strong effect of adjuvant therapy on survival in our cohort. Additional oxaliplatin did not have a marked effect and should be avoided in elderly patients.

ACCOMPANYING CONTENT

 [Data Supplement](#)

Accepted July 31, 2023

Published September 25, 2023

JCO Clin Cancer Inform

7:e2300080

© 2023 by American Society of
Clinical Oncology

Licensed under the Creative
Commons Attribution 4.0 License

INTRODUCTION

With over 8,000 new cases annually, which constitutes approximately 7% of the total cancer incidence, colon cancer is the fourth most common type of cancer in the Netherlands.¹ Given that treatment is always a balance between cost (both financial and physical) and benefit, it is important to continuously assess and improve the effect of (adjuvant) treatment.

For stage III colon cancer, Dutch guidelines recommend treatment with surgery and adjuvant chemotherapy with either capecitabine and oxaliplatin (CAPOX) or infusional

folinic acid-fluorouracil-oxaliplatin (FOLFOX). However, in elderly patients, the benefit of oxaliplatin is still a point of contention.^{2,3}

While 59% of the patients with colon cancer are older than 70 years, elderly patients are frequently excluded from clinical trials on the basis of age alone. Even if age does not rule out participation, elderly patients included may still not be representative for daily clinical practice as they must be fit enough to satisfy other inclusion criteria.

To bridge this gap, observational data play a vital role. Unfortunately, when it comes to estimating treatment

CONTEXT

Key Objective

Adjuvant chemotherapy in elderly is effective although previous analyses dispute the benefit of adding oxaliplatin to fluoropyrimidines. Using observational data to estimate treatment effect always comes with the risk of bias, specifically confounding by indication. Here, we use structure learning algorithms for Bayesian Networks, in conjunction with clinical knowledge, to identify confounders, mitigate this risk, and reliably estimate the effect of adjuvant treatment in colon cancer.

Knowledge Generated

Structure learning can aid in finding causal relationships in observational data by precluding noncausal relationships, thus facilitating identification of confounders. We found a strong effect of adjuvant therapy on survival in elderly patients (70 years and older); additional oxaliplatin provided no further benefit.

Relevance

Identification of and correction for confounders is required for using observational data for estimating treatment effects. The ability to do so greatly enhances the use of observational data. Our research suggests oxaliplatin should be avoided in elderly patients.

effect, observational data are sensitive to bias, specifically selection biases like confounding by indication. For example, if predominantly relatively healthy patients receive treatment, this will skew results in favor of treatment.

There are different ways to account for confounding by indication, for example, through including confounders as covariates (in regression analysis), stratification, or propensity score correction.⁴⁻⁹ In any case, correction requires a decision on which variables should be included as confounders. Several criteria exist to help in this decision, for example, the pretreatment criterion, common cause criterion, backdoor path criterion, and disjunctive cause criterion.¹⁰⁻¹² The pretreatment criterion would select *all* pretreatment variables as confounders. The common cause criterion would, as the name suggests, only correct for those variables thought to be common causes of exposure and outcome. The backdoor path criterion resembles the common cause criterion but takes chains of influence into account and can thus correct for indirect confounders. The disjunctive cause criterion would correct for those variables thought to be either a cause of exposure or outcome.

When there is uncertainty about the presence of unmeasured confounding, it can be reasoned that, generally, the disjunctive cause or backdoor path criteria yield the most unbiased results. For example, correcting for variable C_2 in Figure 1, which would be corrected for when applying the pretreatment criterion, would unnecessarily bias the estimated effect of $T \rightarrow Y$. Application of these criteria, however, requires a causal model. Interestingly, structure learning algorithms for Bayesian Networks (BNs) can be used—with input from domain experts—to discover causal models from data, as recently shown.¹³

Briefly, BNs are a type of probabilistic graphical model whose structure is determined by a directed acyclic graph where

nodes represent variables and directed edges signify (preferably causal) relationships.^{11,14} Each node is associated with a probability distribution conditional on its parents (ie, nodes with a directed edge toward that node).

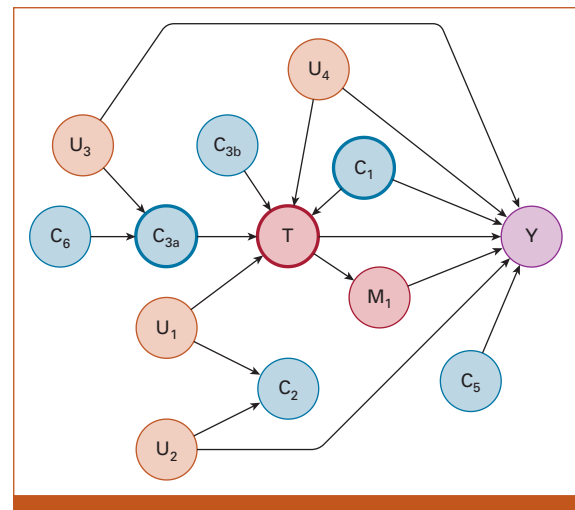


FIG 1. Different types of pathways between treatment (T) and outcome (Y). Blue nodes (C_*) represent measured variables that could be considered potential confounders, and orange nodes (U_*) represent unmeasured variables. Correcting for nodes C_1 and C_{3a} is required to obtain an unbiased result. Correction for C_{3a} and C_6 simultaneously would open the backdoor path through U_3 and thus should be avoided. Correction for unmeasured variables U_1 and U_2 is unnecessary. Adjusting for C_2 would open a backdoor path between T and Y, thus biasing the estimate. Correcting for C_{3b} and/or C_5 would be an unnecessary, bias-neutral adjustment. Correction for M_1 would yield unpredictable results, ranging from overestimating, to nullifying, to reversing any estimated effect. U_4 is the only unmeasured confounder that cannot be corrected for. Figure adapted from VanderWeele and Shpitser¹² (Fig 1).

Structure learning algorithms for BNs can be classified into three categories. Constraint-based algorithms (eg, Necessary Path Condition [NPC]¹⁵) leverage the presence of conditional (in)dependencies to determine a model that best fits the data, score-based algorithms (eg, Silander-Myllymaki [SM]¹⁶) optimize a heuristic that describes goodness of fit, and hybrid algorithms (eg, Max-Min Hill-climbing [MMHC]¹⁷) combine constraint- and score-based strategies.

Generally, these algorithms can incorporate previous (clinical) knowledge in the form of mandatory and prohibited (directed) edges. While causal relations (mandatory edges) between variables may be contested, counterfactuals (prohibited edges) can often be supplied with reasonable certainty and consensus. For example, tumor characteristics, treatment, or outcome will never *cause* age even if the reverse relation (eg, age → tumor characteristics) is contended.

When using structure learning in conjunction with these expert-defined counterfactuals (a blacklist consisting of prohibited edge directions), it is possible to obtain a causal graph. This model can then be used for confounder identification by applying one of the aforementioned criteria and subsequently mitigation.

Previously, we analyzed the effect of adjuvant chemotherapy with CAPOX or capecitabine monotherapy (CapMono) in elderly patients with stage III colon cancer using data from the Netherlands Cancer Registry (NCR) which covers the entire Dutch population.² In this analysis, we adjusted for confounding by indication using the pretreatment criterion.

In this article, we again investigate the treatment effect of adjuvant chemotherapy and the added benefit of oxaliplatin (CAPOX) over CapMono. To this end, we first build causal models using BNs in conjunction with different structure learning algorithms and a varying number of recurrence/survival nodes. Finally, we explore the effect of correcting for different sets of confounders, as identified by these causal models, and compare this with propensity score correction.

METHODS

Data

The original data set from the study by van Erning et al² was used, and vital status was updated until January 2021. The median follow-up for overall survival (OS) was 58 months (after surgery) for all patients and 113 months for patients alive at last follow-up. The median follow-up for recurrence-free survival was 19 months (after surgery) and 29 months for patients alive at last follow-up.

Briefly, a cohort of patients, diagnosed in the southeastern part of the Netherlands between 2005 and 2012 with pathologic stage III (pT₁₋₄N₁₋₂M₀) colon cancer and 70 years and older, was selected from the NCR. The variables sex, age, ASA classification (a system indicating physical performance status,

developed by the American Society of Anesthesiologists), pT, pathological lymph node classification (pN), tumor subsite (coded according to ICD-O-3), and differentiation grade were included. In addition, in 2013 and 2014, details regarding adjuvant therapy, the number of comorbidities, and development of (local) recurrence were acquired from medical records and added to the NCR. The year of diagnosis was available but dropped after a quick analysis of its association with recurrence (Data Supplement, Variable Selection and Fig S1).

As before, patients who died within 90 days after surgery (n = 125) were excluded since these deaths were likely due to surgical complications and patients were unable to undergo adjuvant chemotherapy. Patients receiving chemotherapy other than CAPOX or CapMono (eg, FOLFOX) were excluded. This left a final data set consisting of n = 982 records (Fig 2; Table 1, Data Supplement, Tables S1 and S2). For analysis of CAPOX compared with CapMono, a subset consisting of patients who received adjuvant therapy was created (n = 352).

For assessing association with treatment and for use in the BN, age was discretized into three categories: 70–74, 75–79, and 80 years and older (Data Supplement, Discretization of Age and Fig S2). The anatomic subsite was relabeled as proximal (caecum—splenic flexure of colon), distal (descending—sigmoid colon), or unknown/unspecified.

OS, counted from the date of surgery, was discretized into five Boolean variables representing 1- to 5-year survival, introducing not applicables (NAs; missing values) for patients who had a follow-up of <5 years and were alive at the time of follow-up. The absence of recurrence was discretized similarly. However, because of the limited follow-up, variables representing the absence of recurrence for ≥3 years were skewed toward recurrence or NA. Consequently, these three variables were dropped (Data Supplement, Variable Selection).

Statistical Analyses

BN Development

Using different combinations of outcome variables, full and simplified BNs were developed. The full networks used two nodes/variables to represent recurrence at 1 and 2 years and five nodes for OS at 1, 2, 3, 4, and 5 years. The simplified networks used the last node of each.

The rationale for also evaluating a simplified network is the following. Having multiple outcome nodes at sequential time points might limit the BN ability to find other associations. For example, death at 1 year after diagnosis perfectly predicts death for all subsequent nodes/time points. Therefore, when testing for association between any other variable and a survival node at a later time point, one is essentially testing for association with survival in a specific interval. It can be reasoned that this reduces the (statistical) power to find associations with later time points. For example, if variable 1 is strongly associated with 1-year survival than any,

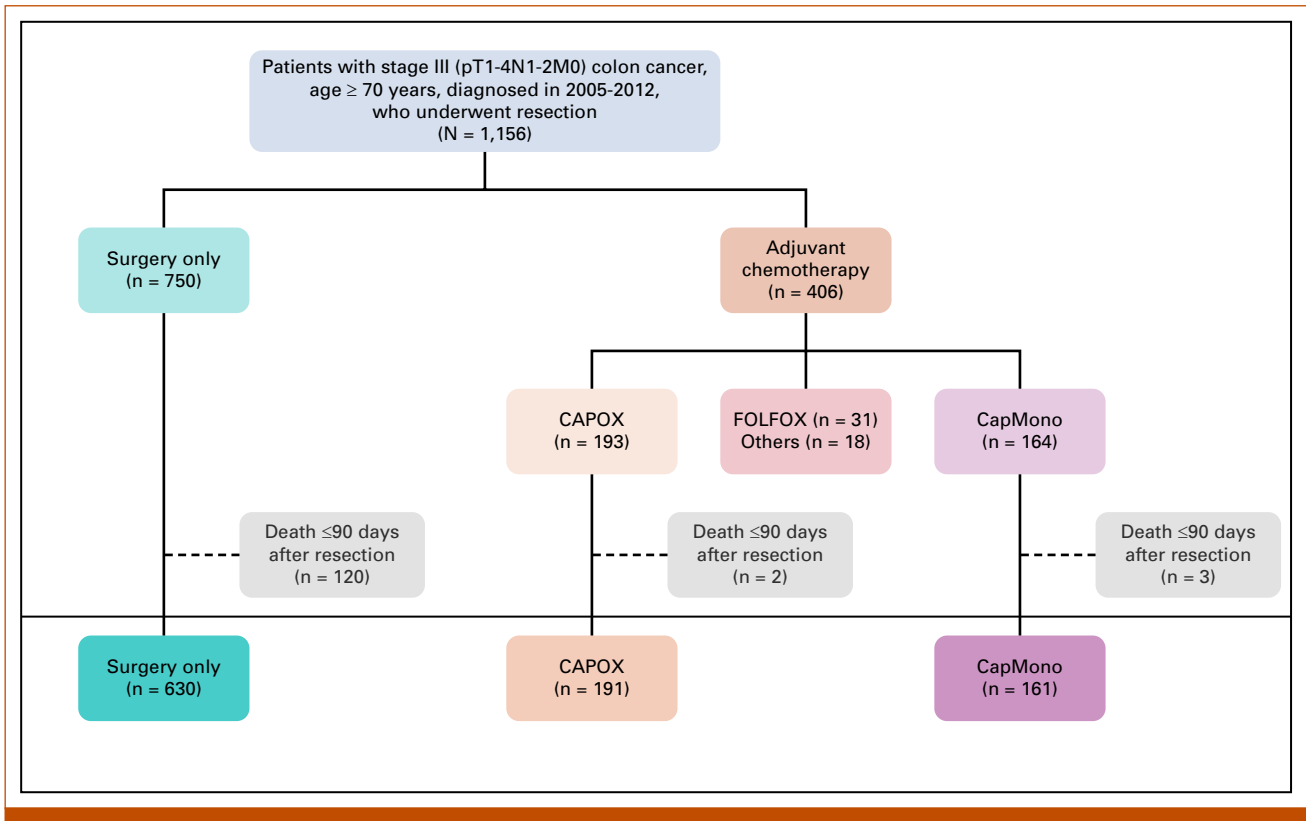


FIG 2. CONSORT diagram—overview of patients included in the study. CapMono, capecitabine monotherapy; CAPOX, capecitabine and oxaliplatin; FOLFOX, infusional fluorouracil, leucovorin, and oxaliplatin.

association of variable 2 with 2-year survival will be harder to detect. In the simplified networks, only the final time point is included to obtain the maximal number of associations with the highest statistical power as the treatment effect on survival is expected to be the highest at that time point.

In addition, we evaluated the effect of using different algorithms than the constraint-based NPC algorithm, as available in Hugin.¹⁸ Specifically, we applied MMHC (hybrid) and SM (score-based) as implemented in the R package bnstruct.^{16,17,19,20} Either adjuvant treatment (Yes, No) or adjuvant treatment regimen (CAPOX, CapMono) was included as appropriate. Obtained BNs were visualized using the R package qgraph.²¹

In all cases, structure learning was performed with the level of significance set to 0.05.¹⁵ In Hugin, the process was supported using a manually defined blacklist of prohibited edge directions (Fig 3). bnstruct uses a different, less granular approach and accepts a list of layers, where edges from lower to higher layers are forbidden.

The NPC algorithm may yield ambiguous regions, consisting of a set of interdependent uncertain links where the absence of one link depends on the presence of another. When this happens, Hugin prompts for input. For resolution, we first

prioritized associations with outcome nodes (recurrence and survival) and subsequently associations with treatment.

In some situations, Hugin ignores the supplied blacklist after manual resolution and misdirects the selected edge. For example, Hugin created an edge from the node representing 4-year survival (surv_04y) toward the ASA node. To overcome this issue, misdirected edges were added to a (minimal) whitelist of enforced edges and structure learning was repeated.

The resulting graphs were inspected to identify confounders.

Propensity Score Analysis

Propensity scores were calculated as follows. First, the chi-squared test was used to assess the association between pretreatment variables and adjuvant treatment. Next, variables with a significant association with treatment were used as covariates in the propensity score calculation. For adjuvant treatment versus surgery only, these were sex, age, comorbidities, ASA, and pN. For CAPOX versus CapMono, these were age and comorbidities only.

Subsequently, the propensity score was discretized using quintiles. Finally, the Cochran-Mantel-Haenszel (CMH) chi-squared test, with the bins as strata, was used to evaluate the effect of propensity score correction.^{8,9,22}

TABLE 1. Variable Distribution and Association Between Variables and Treatment

Variable	Adjuvant Therapy Type			Adjuvant Therapy v Surgery Only				CAPOX v CapMono			
	None	CapMono	CAPOX	P	Sign.	P (CMH)	Sign. (CMH)	P	Sign.	P (CMH)	Sign. (CMH)
Sex, No. (%)											
Male	284 (45.1)	75 (46.6)	108 (56.5)	.044	*	.540		.079		.094	
Female	346 (54.9)	86 (53.4)	83 (43.5)								
Age, years, No. (%)											
70-74	108 (17.1)	52 (32.3)	140 (73.3)	.000	***	.106		.000	***	.129	
75-79	197 (31.3)	84 (52.2)	47 (24.6)								
≥80	325 (51.6)	25 (15.5)	4 (2.1)								
Comorbidities, No. (%)											
None	97 (15.4)	33 (20.5)	65 (34.0)	.000	***	.955		.043	*	.291	
1	135 (21.4)	52 (32.3)	48 (25.1)								
≥2	377 (59.8)	72 (44.7)	74 (38.7)								
Unknown	21 (3.3)	4 (2.5)	4 (2.1)								
ASA, No. (%)											
1	16 (2.5)	12 (7.5)	25 (13.1)	.000	***	.513		.141		.220	
2	241 (38.3)	83 (51.6)	106 (55.5)								
3	220 (34.9)	27 (16.8)	20 (10.5)								
4	4 (0.6)	0 (0.0)	1 (0.5)								
Unknown	149 (23.7)	39 (24.2)	39 (20.4)								
pT, No. (%)											
T1	10 (1.6)	6 (3.7)	4 (2.1)	.259		.157		.286		.400	
T2	51 (8.1)	18 (11.2)	18 (9.4)								
T3	452 (71.7)	107 (66.5)	144 (75.4)								
T4	117 (18.6)	30 (18.6)	25 (13.1)								
pN, No. (%)											
N1	478 (75.9)	107 (66.5)	124 (64.9)	.001	**	.720		.849		.351	
N2	152 (24.1)	54 (33.5)	67 (35.1)								
Location, No. (%)											
Proximal	393 (62.4)	99 (61.5)	98 (51.3)	.142		.991		.159		.428	
Distal	230 (36.5)	60 (37.3)	90 (47.1)								
Others or unknown	7 (1.1)	2 (1.2)	3 (1.6)								
Grade, No. (%)											
g1	34 (5.4)	10 (6.2)	4 (2.1)	.698		.868		.078		.080	
g2	394 (62.5)	95 (59.0)	135 (70.7)								
g3	162 (25.7)	41 (25.5)	41 (21.5)								
g4	1 (0.2)	1 (0.6)	0 (0.0)								
Unknown	39 (6.2)	14 (8.7)	11 (5.8)								

NOTE. For adjuvant therapy versus surgery only, propensity scores were calculated using sex, age, comorbidities, ASA, and pN. For CAPOX versus CapMono, age and comorbidities were used. P: calculated using the Pearson chi-squared test. P* calculated using the CMH Chi-squared test after stratification by propensity score quintile.

Abbreviations: ASA, American Society of Anesthesiologists score; CapMono, capecitabine monotherapy; CAPOX, capecitabine and oxaliplatin; CMH, Cochran-Mantel-Haenszel; pN, pathological lymph node classification; pT, pathological tumor classification; Sign., significance.

*P ≤ .05.

**P ≤ .005.

***P ≤ .0005.

Cox Proportional Hazards Models

Several Cox models were developed, with each investigating the effect of (regimen of) adjuvant treatment on survival

while correcting for a different set of potential confounders, driven by the results from structure learning. Covariates included are listed in Figure 4. Hazard ratios (HRs) for treatment were extracted, together with their CIs.

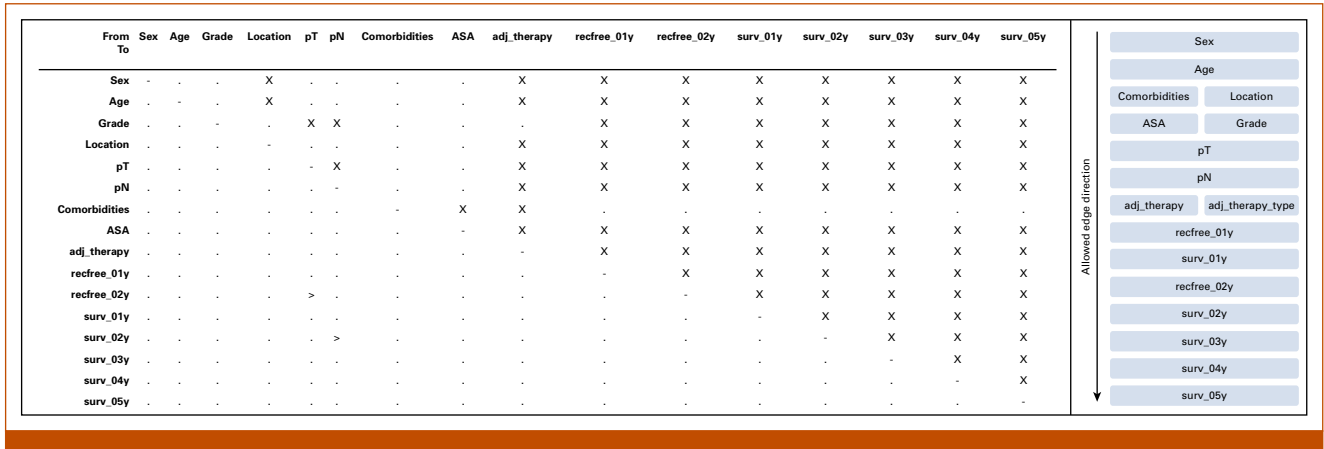


FIG 3. Structure learning constraints for the effect of adjuvant chemotherapy on overall survival. Left: constraints for Hugin. The blacklist is made up from cells containing x: these edges are forbidden. Minimal whitelist (ie, enforced edges) is formed by cells marked with >. Right: constraints for bnstruct. Edges can only originate from a node in a layer at the same height or higher. For example, sex → age is allowed, but not the reverse. Edges between nodes at the same height can point either way. ASA, American Society of Anesthesiologists score; pN, pathological lymph node classification; pT, pathological tumor classification.

RESULTS

Adjuvant Treatment Versus Surgery Only

Variable Distribution and Propensity Score Correction

Chi-squared analysis showed that variables age, sex, pN, ASA (physical performance status), and comorbidities were significantly associated with receiving any form of adjuvant treatment (Table 1; Data Supplement, Table S1). Stratification using the discretized propensity score as strata in the CMH chi-squared test removed all associations between these conceivable confounders and treatment.

BN Development and Confounder Identification

The full BNs are shown in the left column of Figure 5. In the BN found by the NPC algorithm, two confounders were identified through inspection of the graph: pN and ASA. The BNs found by MMHC and SM algorithms did not identify any confounders and generally had fewer edges.

In the simplified networks, which focused on a limited set of recurrence and survival nodes (Fig 5, right column), the NPC algorithm again identified pN as confounder, but ASA, which was associated with surv_04y in the full network, was no longer associated with survival. The SM algorithm

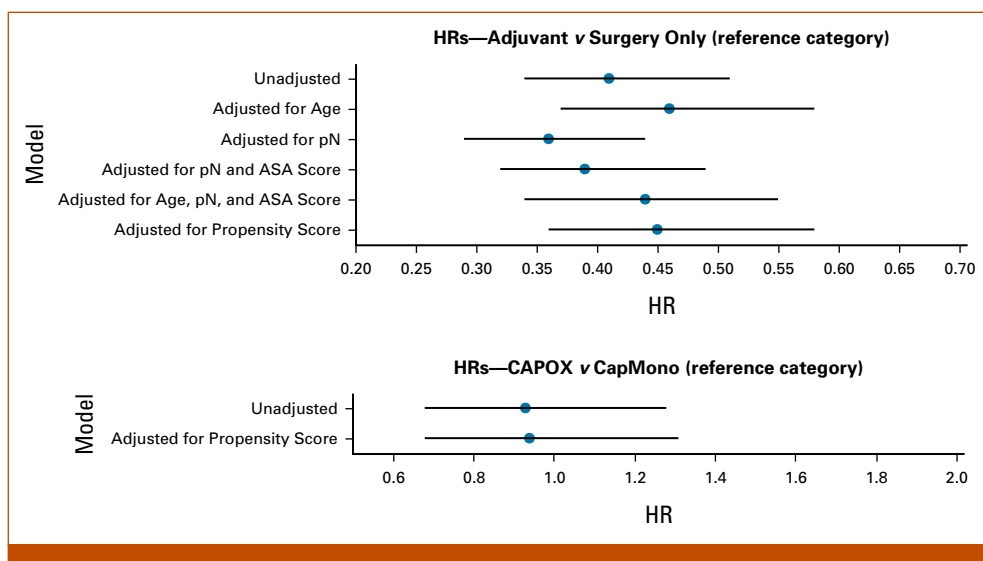


FIG 4. HRs (with CIs). Top: adjuvant treatment versus surgery only (referent category). Bottom: CAPOX versus CapMono (referent category). ASA, American Society of Anesthesiologists score; CapMono, capecitabine monotherapy; CAPOX, capecitabine and oxalipatin; HR, hazard ratio; pN, pathological lymph node classification.



FIG 5. Bayesian Networks for adjuvant treatment versus surgery only. Rows 1-3 correspond to the algorithms NPC, MMHC, and SM, respectively. The left column shows full networks, and the right column shows simplified networks. Treatment indicates whether any form of adjuvant treatment was received. Top left: pN and ASA act as confounders. Top right: pN acts as a confounder. Middle left: no confounders identified; pN, ASA, and treatment are independent. Middle right: no confounders identified. Bottom left: age acts as a confounder. ASA, American Society of Anesthesiologists score; MMHC, Max-min Hill-climbing; NPC, Necessary Path Condition; pN, pathological lymph node classification; pT, pathological tumor classification; SM, Silander-Myllymaki.

did not find an association between pN and treatment (node adj_therapy) but *did* find an association between age and surv_05y, identifying age as a potential confounder.

In summary, depending on the algorithm selected and nodes included, treatment was either unconfounded or confounded by pN, pN and ASA, or age.

Cox Proportional Hazards Models

Adjuvant treatment was significantly associated with survival in all models (Fig 4, top row; Data Supplement, Table S3). Estimated HRs were between 0.39 and 0.45, depending on covariates included. The models had overlapping CIs and yielded comparable results.

CAPOX Versus CapMono

Variable Distribution and Propensity Score Correction

Chi-squared analysis showed variables age and comorbidities to be significantly associated with the choice between CAPOX or CapMono (Table 1; Data Supplement, Table S2). Here too, stratification using the binned propensity score as strata in the CMH chi-squared test removed all associations between these suspected confounders and treatment.

BN Development and Confounder Identification

From the full networks, none of the algorithms identified an association between the adjuvant treatment regimen and survival (Fig 6). From the pretreatment variables, only grade was associated with survival (using the NPC algorithm) and no confounders were found.

In the simplified networks, again, none of the algorithms identified an association between the adjuvant treatment

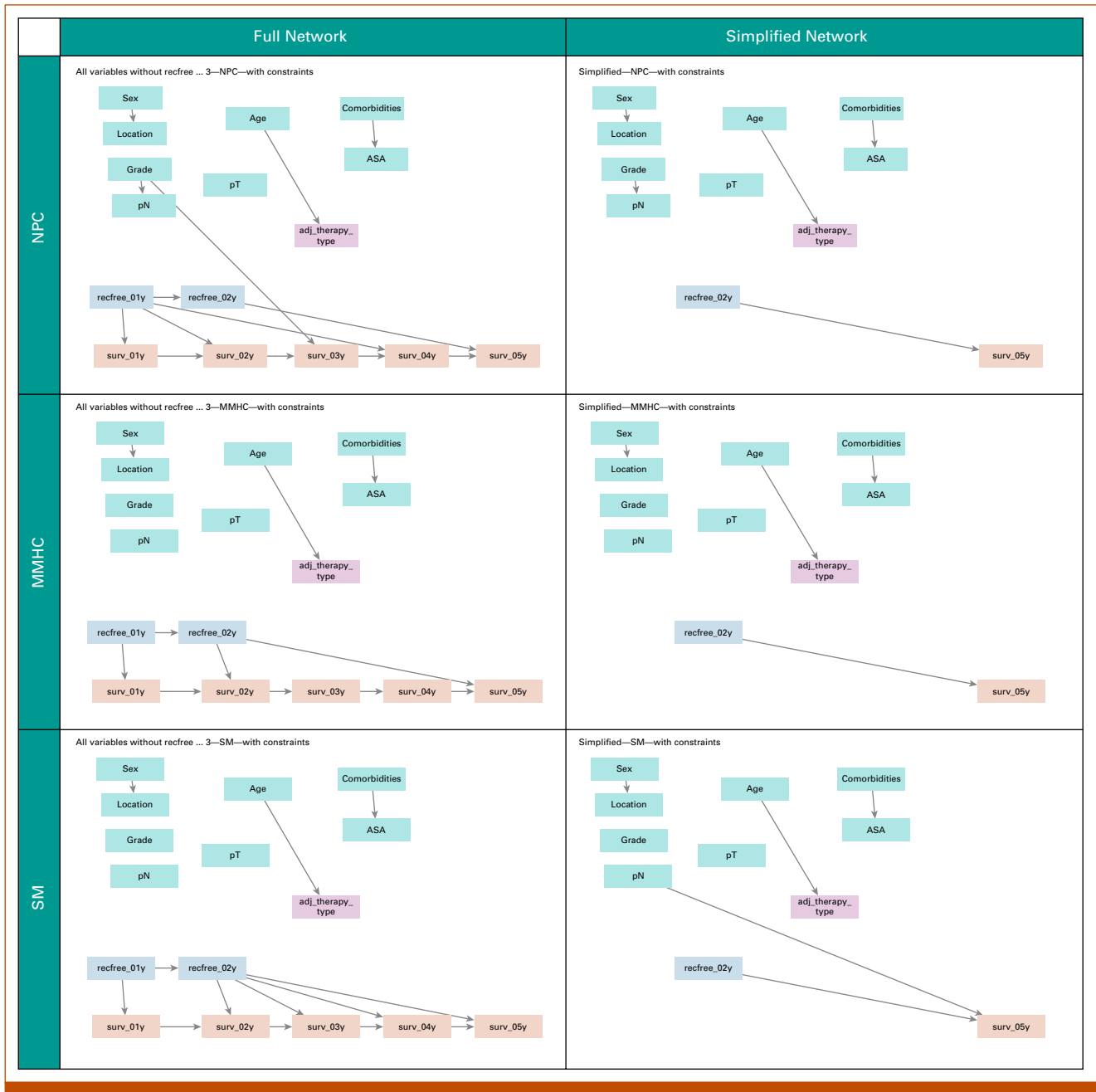


FIG 6. Bayesian Networks for CAPOX versus CapMono. Rows 1-3 correspond to the algorithms NPC, MMHC, and SM, respectively. The left column shows full networks, and the right column shows simplified networks. Treatment indicates whether CAPOX or CapMono was received. Top left: no confounders found. Top right: no confounders found. Middle left: no confounders found. Middle right: no confounders found. Bottom left: no confounders found. Bottom right: no confounders found. ASA, American Society of Anesthesiologists score; CapMono, capecitabine monotherapy; CAPOX, capecitabine and oxaliplatin; MMHC, Max-min Hill-climbing; NPC, Necessary Path Condition; pN, pathological lymph node classification; pT, pathological tumor classification; SM, Silander-Myllymaki.

regimen and survival. All networks suggested that age was a driving factor for the choice of regimen. The SM algorithm implied that pN was an independent predictor of survival.

To summarize, regardless of the algorithm selected and nodes included, choice of treatment regimen was not only unconfounded but also unassociated with survival.

Cox Proportional Hazards Models

None of the Cox models found significant association between the adjuvant treatment regimen and survival with HR point estimates around 0.93 and CIs equal to or wider than 0.68–1.28 (Fig 4, bottom row; Data Supplement, Table S4). CIs between the models were largely overlapping.

DISCUSSION

We were able to successfully leverage BN structure learning algorithms in conjunction with (basic) clinical knowledge to create causal models and subsequently identify potential confounders. Different structure learning algorithms identified different potential confounders, but none were contradicting. In other words, confounders identified in one model were not mediators or colliders in others (Data Supplement, Fig S3), which may introduce bias if corrected for. From a theoretical point of view, score-based algorithms run a higher risk of incorrectly classifying a variable as a confounder in the presence of *unmeasured* confounding (Data Supplement, Fig S4). In this case, this mechanism does not seem to play a significant role since all HRs were very close together.

When comparing adjuvant treatment with surgery only, the set of identified confounders differed, depending on both the specific algorithm used and the nodes included in the network. Potential confounders identified were pN, ASA, and age although 3 (out of 6) networks suggested that the treatment effect was unconfounded. All Cox models found significant benefit of adjuvant treatment over surgery only (HRs ranging between 0.36 and 0.46); correcting for different sets of potential confounders had no marked effect, and neither did correction using propensity score.

In the comparison between CAPOX and CapMono, no confounders were identified, regardless of the algorithm or nodes included: none of the BNs found an association between the choice of treatment regimen and survival. The unadjusted Cox model yielded the same conclusion and did not find any effect of adding oxaliplatin. Correcting for propensity score did not make a difference.

Looking at the BNs in general, several interesting associations between pretreatment variables are found in all networks, regardless of algorithm or node selection. Unsurprisingly, age appears to be associated with the choice of adjuvant treatment, not only when considering whether to start either regimen but also when making a choice for a

specific regimen. Similarly, sex was associated with the tumor location and grade with pN, both associations that have been reported before.^{23,24} The association found between the number of comorbidities and ASA score also makes sense considering the use of the ASA score (physical performance status) to approximate comorbidity in registries.²⁵ On the other hand, some expected associations were not, or only infrequently, found: only one network connected pT and survival. However, this might be explained by the fact that we selected stage III and the (vast) majority of patients had a pT3 tumor. Overall, we considered the obtained networks clinically plausible.

Previously, Wilkinson et al²⁶ investigated the effect of adjuvant 5-fluorouracil with leucovorin compared with surgery only. The HR found here is smaller than the effect they reported, which was 0.62 in stage II–III and 0.65 in stage III only. It should be noted that the age distribution between our study (100% for patients older than 70 years) and the study by Wilkinson et al (approximately 36%–51% patients younger than 60 years depending on the treatment group) is vastly different. Our results seem to suggest that, in our elderly cohort, chemotherapy improves survival relatively more.

The effect of adding oxaliplatin to a fluoropyrimidine (eg, capecitabine or fluorouracil) in stage II–III colon cancer has been extensively studied.^{3,27–32} Three trials in particular have contributed to our understanding: the NO16968 (XELOX in Adjuvant Colon Cancer Treatment) study, the Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) trial, and the C-07 trial of the National Surgical Adjuvant Breast and Bowel Project (NSABP).

NO16968 and MOSAIC found a small (4–6 percentage point) but significant benefit of adding oxaliplatin to the treatment regimen. In MOSAIC, treatment effect was stronger in patients with stage III compared with stage II; NO16968 only targeted stage III patients.^{29,33} Both trials included a relatively young population. NSABP, which also investigated a younger population compared with our analysis, did not find an overall benefit of adding oxaliplatin but did report a significant effect in an unplanned subset analysis of patients 70 years and younger.³² In a subgroup analysis, the MOSAIC study drew the complimentary conclusion that there is no additional benefit of oxaliplatin in elderly patients. This is in line with the results obtained here.

In conclusion, we have shown that structure learning elucidates underlying relationships in data, helping select which variables should be corrected for. Varying included variables and algorithms yielded (slightly) different, complementary results and identified different sets of potential confounders. HRs were similar regardless of the chosen set.

We found a strong association between adjuvant treatment with capecitabine and survival in stage III colon cancer in our

cohort of patients 70 years and older. No additional benefit of adding oxaliplatin was found. As such, addition of oxaliplatin

may be considered in younger patients with more advanced stage but should be avoided in elderly patients.

AFFILIATIONS

¹Department of Research and Development, Netherlands Comprehensive Cancer Organization, Utrecht, the Netherlands

²Department of Radiation Oncology (Maastr), GROW School for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, the Netherlands

³Department for Health Evidence, Radboud University Medical Centre, Nijmegen, the Netherlands

⁴Department of Surgery, Catharina Hospital, Eindhoven, the Netherlands

CORRESPONDING AUTHOR

Melle Sieswerda, MD, Department of Research and Development, Netherlands Comprehensive Cancer Organisation, Godebaldkwartier 419, Utrecht 3511 DT, the Netherlands; Twitter: @mellesies; e-mail: m.sieswerda@iknl.nl.

DATA SHARING STATEMENT

Source code accompanying this manuscript can be found online at <https://github.com/mellesies/Using-BNs-for-estimating-treatment-effect-in-CRC>.

AUTHOR CONTRIBUTIONS

Conception and design: Melle Sieswerda, Ruby van Rossum, Inigo Bermejo, Gijs Geleijnse, Valery Lemmens, André Dekker, Xander Verbeek

Administrative support: Inigo Bermejo

Provision of study materials or patients: Felice van Erning

Collection and assembly of data: Ruby van Rossum, Felice van Erning

Data analysis and interpretation: Melle Sieswerda, Ruby van Rossum, Inigo Bermejo, Katja Aben, Felice van Erning, Ignace de Hingh, André Dekker

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted.

I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information

about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](http://OpenPayments)).

Gijs Geleijnse

Research Funding: AstraZeneca (Inst), Roche (Inst), Janssen (Inst)

Katja Aben

Research Funding: Netherlands Comprehensive Cancer Organisation (Inst)

Ignace de Hingh

Consulting or Advisory Role: Alesi Surgical Limited (Inst)

Research Funding: Roche (Inst), QP&S/RanD (Inst)

André Dekker

Employment: Medical Data Works B.V

Stock and Other Ownership Interests: Medical Data Works B.V

Honoraria: Varian Medical Systems, Accuray, Roche, Janssen-Cilag, Philips Research (Inst), Varian Medical Systems (Inst), Medtronic

Consulting or Advisory Role: Varian Medical Systems

Research Funding: Varian Medical Systems (Inst), Philips Healthcare (Inst), OncoRadiomics (Inst)

Patents, Royalties, Other Intellectual Property: Royalties from Mirada Medical Ltd on a deep learning application in medical imaging (Inst), Royalties from Varian Medical Systems on dosimetry software for radiation oncology (Inst), Royalties from Health Innovation Ventures on Radiomics, eLearning, and prediction models in cancer (Inst), Royalties from PXI for software related to small animal irradiation for life sciences research (Inst), Patents held currently—Monitoring respiration based on plethysmographic heart rate signal. US patent 6,702,752—Apparatus and method for monitoring respiration with a pulse oximeter. US patent 6,709,402—Monitoring Mayer wave effects based on a photoplethysmographic signal. US patent 6,805,673—Monitoring physiological parameters based on variations in a photoplethysmographic baseline signal. US patent 6,896,661—Monitoring physiological parameters based on variations in a photoplethysmographic signal. US patent 7,001,337—Knowledge-Based Interpretable Predictive Model for Survival Analysis. US Patent 8,078,554—Dose distribution modeling by region from functional imaging. US Patent 8,812,240—Systems, methods and devices for analyzing quantitative information obtained from radiological images US Patent 9721340 B2

Expert Testimony: Thoratec

Travel, Accommodations, Expenses: Medical Data Works B.V, Varian Medical Systems, Roche, Janssen-Cilag

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

The authors thank the registration team of the Netherlands Comprehensive Cancer Organisation (IKNL) for the collection of data for the Netherlands Cancer Registry.

REFERENCES

1. NKR Cijfers: Incidentie—Grafiek, 2023. <https://nkr-cijfers.iknl.nl>
2. van Erning F, Janssen-Heijnen MLG, Creemers GJ, et al: Recurrence-free and overall survival among elderly stage III colon cancer patients treated with CAPOX or capecitabine monotherapy. *Int J Cancer* 140:224-233, 2017
3. Tournigand C, André T, Bonnetain F, et al: Adjuvant therapy with fluorouracil and oxaliplatin in stage II and elderly patients (between ages 70 and 75 years) with colon cancer: Subgroup analyses of the Multicenter International Study of Oxaliplatin, Fluorouracil, and Leucovorin in the Adjuvant Treatment of Colon Cancer trial. *J Clin Oncol* 30:3353-3360, 2012
4. Rubin DB, Thomas N: Matching using estimated propensity scores: Relating theory to practice. *Biometrics* 52:249-264, 1996
5. Rubin DB: Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 127:757-763, 1997
6. Li J, Handorf E, Bekelman J, et al: Propensity score and doubly robust methods for estimating the effect of treatment on censored cost. *Stat Med* 35:1985-1999, 2016
7. Genbäck M, de Luna X: Causal inference accounting for unobserved confounding after outcome regression and doubly robust estimation. *Biometrics* 75:506-515, 2019
8. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22:719-748, 1959
9. Mantel N: Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *J Am Stat Assoc* 58:690-700, 1963

10. Häggström J: Data-driven confounder selection via Markov and Bayesian Networks. *Biometrics* 74:389-398, 2018
 11. Pearl J: *Causality: Models, Reasoning and Inference* (ed 2). New York, NY, Cambridge University Press, 2009
 12. VanderWeele TJ, Shpitser I: A new criterion for confounder selection. *Biometrics* 67:1406-1413, 2011
 13. Sieswerda M, Xie S, van Rossum R, et al: Identifying confounders using Bayesian Networks and estimating treatment effect in prostate cancer with observational data. *JCO Clin Cancer Inform* 7:e2200080, 2023
 14. Koller D, Friedman N: *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, Massachusetts; London, England, MIT Press, 2009
 15. Steck H: Constraint-based structural learning in Bayesian Networks using finite data sets [dissertation]. Munich Technical University, Munich, Germany, 2001
 16. Silander T, Myllymaki P: A simple approach for finding the globally optimal Bayesian Network structure, 2012. <http://arxiv.org/abs/1206.6875>
 17. Tsamardinos I, Brown LE, Aliferis CF: The max-min hill-climbing Bayesian Network structure learning algorithm. *Mach Learn* 65:31-78, 2006
 18. Andersen SK, Olesen KG, Jensen FV, et al: HUGIN—A shell for building Bayesian Belief universes for expert systems. *Proc Elev Int Jt Conf Artif Intell* 2:1080-1085, 1989
 19. R Core Team: *R: A language and environment for statistical computing*. Vienna, Austria, 2022. <https://www.R-project.org/>
 20. Franzin A, Sambo F, di Camillo B: bnstruct: an R package for Bayesian Network structure learning in the presence of missing data. *Bioinformatics* 33:1250-1252, 2017
 21. Epskamp S, Cramer AOJ, Waldorp LJ, et al: qgraph: Network visualizations of relationships in psychometric data. *J Stat Softw* 48:1-18, 2012
 22. Agresti A: *Categorical Data Analysis* (ed 2). New York, NY, Wiley, 2002
 23. DeCosse JJ, Ngoi SS, Jacobson JS, et al: Gender and colorectal cancer. *Eur J Cancer Prev* 2:105-116, 1993
 24. Marija C, Kresimir D, Ognjen B, et al: Estimation of colon cancer grade and metastatic lymph node involvement using DWI/ADC sequences. *Acta Radiol* 64:1341-1346, 2022
 25. Quach LH, Jayamaha S, Whitehouse SL, et al: Comparison of the Charlson Comorbidity Index with the ASA score for predicting 12-month mortality in acute hip fracture. *Injury* 51:1004-1010, 2020
 26. Wilkinson NW, Yothers G, Lopa S, et al: Long-term survival results of surgery alone versus surgery plus 5-fluorouracil and leucovorin for stage II and stage III colon cancer: Pooled analysis of NSABP C-01 through C-05. A baseline from which to compare modern adjuvant trials. *Ann Surg Oncol* 17:959-966, 2010
 27. Haller DG, Taberero J, Maroun J, et al: Capecitabine plus oxaliplatin compared with fluorouracil and folinic acid as adjuvant therapy for stage III colon cancer. *J Clin Oncol* 29:1465-1471, 2011
 28. André T, Boni C, Mounedji-Boudiaf L, et al: Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med* 350:2343-2351, 2004
 29. André T, de Gramont A, Vernerey D, et al: Adjuvant fluorouracil, leucovorin, and oxaliplatin in stage II to III colon cancer: Updated 10-year survival and outcomes according to BRAF mutation and mismatch repair status of the MOSAIC study. *J Clin Oncol* 33:4176-4187, 2015
 30. Aparicio T, Francois E, Cristol-Dalstein L, et al: PRODIGE 34-FFCD 1402-ADAGE: Adjuvant chemotherapy in elderly patients with resected stage III colon cancer: A randomized phase 3 trial. *Dig Liver Dis* 48:206-207, 2016
 31. Aparicio T, Bouché O, Etienne P-L, et al: Preliminary tolerance analysis of adjuvant chemotherapy in older patients after resection of stage III colon cancer from the PRODIGE 34-FFCD randomized trial. *Dig Liver Dis* 55:541-548, 2022
 32. Yothers G, O'Connell MJ, Allegra CJ, et al: Oxaliplatin as adjuvant therapy for colon cancer: Updated results of NSABP C-07 trial, including survival and subset analyses. *J Clin Oncol* 29:3768-3774, 2011
 33. Schmoll H-J, Taberero J, Maroun J, et al: Capecitabine plus oxaliplatin compared with fluorouracil/folinic acid as adjuvant therapy for stage III colon cancer: Final results of the N016968 randomized controlled phase III trial. *J Clin Oncol* 33:3733-3740, 2015
-