Adopting ^{18}F-FDG PET/CT in routine first-line follow-up after thermal ablation therapy in patients with resectable colorectal liver metastases – An explorative cost-effectiveness analysis based on clinical trial data

Sietse van Mossel

Industrial Engineering and Management, Health Care Technology and Management

Department of Health Technology & Services Research, University of Twente

Department of Radiology, Leiden University Medical Center

Department of Surgery, Leiden University Medical Center

Graduation Manuscript

June 27, 2022

Summary

Introduction: This comparative study was conducted to assess, in terms of cost-effectiveness, the clinical benefits of adopting ^{18}F-FDG PET/CT to conventional contrast-enhanced CT (CECT) imaging for patients with oligometastatic colorectal liver metastases (CRLM) after receiving thermal ablation (TA) therapy in first-line follow-up. It was investigated whether the routine use of ^{18}F-FDG PET/CT in first-line follow-up cost-effectively improves health outcomes through improved clinical decision-making and patient management.

Methods: The study design was twofold. First, we conducted a single-centre multidisciplinary clinical trial complemented by retrospectively searching the hospital's electronic patient database. Second, we conducted a treatment-driven discrete event simulation (DES). Individual patient data collected in the clinical trial were analysed to determine accurate input parameter values for the DES study. For each input parameter, a distribution describing the variation at patient-level was defined. A literature search was performed to collect aggregated evidence on health outcomes and CRLM care pathway-related costs. To enhance the quality of secondary survival data analyses, we applied the algorithm of *Guyot et al.* which derives from published Kaplan-Meier curves a close approximation to the original individual patient-level time-to-event data from which they were generated. Incremental cost-effectiveness ratios (ICERs) were calculated by dividing the estimated mean difference in costs by the mean difference in quality-adjusted life years (QALYs) or life years gained (LYGs). The consequences of uncertainty in input parameter values on model outcomes were assessed through probabilistic analysis (PA).

Results: Through PA, we simulated a minimum of 10,000 patients per run which was a sufficiently large number to remove the impact of patient-level variation in our cost-effectiveness outcomes. When performing 10,000 runs, the mean number of cases of residual/local recurrence observed in first-line follow-up was nearly doubled when adopting ^{18}F-FDG PET/CT in first-line follow-up (1530 versus 888 cases). Furthermore, the mean number of false-negative test results in first-line follow-up was nearly halved (737 versus 1374 cases). Consecutively, we observed a small increase in the number of consecutive LAT procedures provided when adopting ^{18}F-FDG PET/CT (4859 versus 4755 cases). Consequently, we could waive (or postpone) providing less favourable complex systemic therapy in a meaningful number of patients. Finally, the summarised lifetime health and economic consequences in the case of adopting ^{18}F-FDG PET/CT resulted in an ICER of €17,850.12 saved per QALY lost, which was a non-dominant ratio, and, €14,668.67 saved per LYG, which was a dominant ratio, respectively.

Conclusions: Adopting ^{18}F-FDG PET/CT in first-line follow-up was not proven to be a cost-effective alternative to conventional follow-up by CECT. However, note that conclusions about our findings regarding lifetime health and economic outcomes should be drawn with caution, as the results of our scenario analyses did not prove to be robust.

Discussion: Our health economic results are in line with previously published cost-effectiveness studies. It is suggested that the use of ^{18}F-FDG PET/CT may lead to a change in the originally planned patient management in up to 20% of all CRLM patients. However, the quality and quantity of available observational data resulted in a large uncertainty surrounding the parameter values used for simulation modelling. Finally, we argue that this study could potentially influence daily clinical practice but further research is needed to determine the cost-effectiveness when integrating multiple lines of consecutive treatment strategies.

Keywords: Cost-effectiveness, colorectal liver metastases, follow-up, ^{18}F-FDG PET/CT, thermal ablation therapy, discrete event simulation, treatment-driven

Introduction

On a global level, colorectal cancer (CRC) introduces a substantial health burden on patients and societies due to its relatively high incidence and mortality. In 2020, CRC was the third most commonly diagnosed cancer accounting for about 10% of all new cancer cases. Moreover, CRC was the second leading cause of cancer mortality and accounted for 9.4% of all cancer-related deaths in the world. (1) Approximately 15-25% of all the patients with diagnosed CRC history suffer metastatic CRC disease (mCRC). About 20% of these patients present with synchronous metastases and up to 60% develop metachronous metastases within three years after primary diagnosis. (2–4)

After apparently curative resection of the primary colorectal tumour, the liver is the most common and first metastatic site of CRC (2–11). In approximately half of the patients with oligometastatic colorectal liver metastases (CRLM), the metastases occur in the liver only, in which curative intended local aggressive treatment (LAT) reaches a 5-year overall survival of approximately 50-60%. However, due to several criteria such as the tumour volume, lack of future liver remnant, location of metastases near vital structures, multifocality of the disease and a poor clinical condition of the patient, only a minority group of patients is eligible for LAT. Neoadjuvant therapy could shrink metastases sufficiently to enable LAT for a subset of patients with initially unresectable CRLM, but for the majority of patients, the aim remains to prolong survival and maintain quality of life by receiving systemic therapy. (2-4,12–19)

CRLM recurrence following LAT is prevalent. Approximately 90% of all recurrences are diagnosed within the first three years after LAT was received. Curation highly depends on the advancement of recurrent disease and is determined in multidisciplinary review board meetings. During these meetings, prognostic factors are proposed to achieve accurate overall survival prediction and to restrict LAT to those patients who will strictly benefit. (20,21) Several papers assessed the most relevant prognostic tumour and patient characteristics; no prognostic and predictive factors were common in all models, though there was a tendency towards the age and clinical condition of patients, the primary CRC site, the primary tumour stage, the mutational status, the disease-free duration, synchronous or metachronous metastases, the number of metastases, the spread to lymph nodes, the maximum size of metastases, bilateral disease, CEA level and extrahepatic spread as representing meaningful independent risk factors (11,17,21–40). Consequently, because of the lack of clarity, several studies aimed to establish a comprehensive prognostic scoring system to achieve more accurate overall survival prediction and proper patient selection. The scoring system should be unambiguous, based on established prognostic patient and disease characteristics, and should not require additional diagnostic testing. (35,41–51)

Surgical metastasectomy of tumour lesions is considered the golden standard and is the most provided LAT option for patients suffering from resectable CRLM. Nonetheless, repeat surgical metastasectomy can be challenging due to adhesions and reduced liver volume (29,34,52–54). Thermal ablation (TA)

could be considered a valid, less invasive and attractive alternative to metastasectomy. Minimally invasive TA techniques lead to lower blood loss, lower complication rates, a shorter duration of hospital stay and lower incremental costs. Also, iterative TA therapies are supported to be safe and effective, because it is associated with health outcomes similar to first-line TA. Particularly in elderly and patients with severe comorbidity, TA is considered an effective alternative LAT option. (7,9,22,52,55–62)

A variety of TA approaches has evolved to complement surgical metastasectomy, or as an autonomous treatment modality, for otherwise unresectable CRLM. Radiofrequency ablation (RFA) and microwave ablation (MWA) are currently the most frequently provided TA techniques (7,29,63,64). Evidence suggests that both TA techniques present similar complications, disease-free survival and overall survival rates (7,65,66). Compared to RFA, however, it is suggested that MWA improves local tumour control while significantly shortening the operative time (67). On the other hand, both TA techniques have lower accuracy and show a higher rate of local/residual recurrence than surgical metastasectomy. Therefore, evolutions in the imaging field during post-operative follow-up workup including ^{18}F-FDG PET/CT are essential to reach increased efficacy rates of TA therapy. (15,16,66,68–70)

The main aims of the diagnostic follow-up following LAT are early and asymptomatic detection of residual tumours and local tumour progression **(1)**, and, the detection of new intrahepatic non-local and extrahepatic distant metastases **(2)**. Anatomical imaging by conventional contrast-enhanced CT (CECT) plus serum carcinoembryonic antigen level (CEA) is considered the golden standard and is, traditionally, the most provided diagnostic for patients suffering from resectable CRLM. MRI abdomen with Primovist is often used as an additional modality in case of suspected tumour recurrence.

The adoption of integrated ^{18}F-FDG PET/CT combines anatomical and metabolic imaging. ^{18}F-FDG PET/CT provides complementary metabolic information that enables the detection of malignant disease at unexpected sites or in morphologically normal structures that may be easily overlooked on cross-sectional imaging. So that early detection of CRLM and tumour activity is demonstrated before structural tissue changes become detectable. Also, purely structural tissue changes can be misleading and do not always reflect tumour aggressiveness correctly. (17,71,72) Moreover, the effectiveness of ^{18}F-FDG PET/CT is emphasized in patients with consecutively elevated CEA levels or patients with potential false-negative results on conventional imaging, as elevated levels of CEA cannot provide accurate localization to a potential site of recurrence (10,73).

However, the precise role of ^{18}F-FDG PET/CT imaging in current national guidelines remains controversial. During imaging assessment, the main challenge appears to distinguish residual tumours from false-positive signs due to postinterventional non-malignant changes such as inflammation or necrosis in the ablation zone (74). The most characteristic advantage of the ^{18}F-FDG PET/CT is the ability to quantify tumour biology using FDG uptake by measuring the standardized uptake value (75,76). Accordingly, several studies quantitatively show a clinical superiority of ^{18}F-FDG PET/CT over

CECT in detecting local tumour progression after TA therapy. It is indicated that the use of ^{18}F-FDG PET/CT in this setting particularly allows for minimization of the false-negative rate compared with CECT, without compromising the low false-positive rate. (77–85)

This comparative study is performed to assess, in terms of cost-effectiveness, the clinical advantages of adopting ^{18}F-FDG PET/CT imaging to first-line follow-up by conventional CECT for patients with CRLM after receiving TA therapy. It explores whether routinely adopting ^{18}F-FDG PET/CT in first-line follow-up cost-effectively improves health outcomes through enhanced clinical decision-making and patient management.

Methods

The study design was twofold. First, we conducted a single-centre controlled clinical trial. Ethical approval was waived since it was considered not obligatory under Dutch law as this study provides an anonymized dataset (determined by the METC LDD with reference no. G21.061). Informed consent was given by patients and obtained by the clinician before reporting data to the cancer registry. All methods were carried out following the clinical guidelines and regulations.

Second, we performed a discrete event simulation (DES) to study the cost-effectiveness of adopting ^{18}F-FDG PET/CT imaging to routine first-line follow-up. The individual patient data gathered in the clinical trial were analysed to define accurate input parameter values for the DES study. For each input parameter, a distribution was defined that describes the variation at patient-level (86).

This study was reported according to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist (87). Also, a recently published verification checklist to reduce errors in models and improve their credibility (TECH-VER) was applied during our study (88).

1 Data gathering process

Data were prospectively collected from patients with resectable CRLM receiving LAT at Leiden University Medical Centre (LUMC) from 01-06-2020 until 31-01-2022. Within the LUMC the concerned departments were the departments of Radiology (section of Nuclear Medicine, Intervention Radiology and Abdominal Radiology) and Surgery. When required, additional recollecting of data was performed by retrospectively searching the hospital's electronic patient database. In addition, a literature search was performed to collect aggregated evidence on health outcomes and CRLM care pathway-related costs. Leftover evidence gaps were addressed by an extended literature search and multi-centre expert elicitation. The single-centre LUMC database, expert elicitation and literature search, together, enclosed a state-of-the-art starting point for explorative health economic evaluation.

2 Study population

Patients were included in this study if they underwent open or minimally invasive metastasectomy, RFA and/or MWA or a combination of these treatments; and an abdominal ^{18}F-FDG PET/CT was made within 6 months following LAT. Patients were excluded when no LAT was performed or if LAT was followed by adjuvant chemotherapy affecting the metabolic activity of tumour cells, and when no ^{18}F-FDG PET/CT was made within 6 months of follow-up. An overview of data on the clinical condition of the patient, TNM-classification of the primary tumour, location of primary tumour and hepatic metastases characteristics, history of systemic therapy, Fong clinical risk score, and previous local hepatic treatments provided, were available via the hospital's electronic patient database.

Demographics

A total of 45 patients with unique LAT procedures were prospectively included. Each unique LAT procedure was followed by imaging with ^{18}F-FDG PET/CT and a conventional CECT scan. Patient characteristics (age, clinical condition), tumour characteristics (primary CRC site, recurrent disease, clinical risk score) and treatment characteristics (induction therapy received, treatment type received) are summarised in Table 1.

Table 1 Demographics of included patients in clinical trial

Study population (n=45)	**Value (%)**
Patient's mean age at diagnosis	64.62 (SD=11.63)
Number of unique LAT procedures	45
Number of unique scans assessed	90
^{18}F-FDG PET/CT	45/90
Conventional CECT	45/90
Site of primary CRC	
Colon	30/45 (67)
Rectum	15/45 (33)
Clinical risk score [1]	
Fong ≥ 3, "high risk"	18/45 (40)
Fong < 3, "low risk"	14/45 (31)
Missing values	13/45 (29)
Clinical condition [2]	
ASA ≥ 3, "poor condition"	15/45 (33)
ASA < 3, "good condition"	30/45 (67)
Recurrent disease	
Yes	23/45 (51)
No	22/45 (49)
Induction therapy received	
Yes	13/45 (29)
No	32/45 (71)
Treatment type received	
Autonomous TA therapy	34/45 (76)
TA additional to metastasectomy	11/45 (24)

[1] The *Fong Clinical Risk Score* for CRC recurrence assigns CRC patients with liver metastases a score of 0-5 based on five independent preoperative risk factors to estimate the 5-year survival and median months of survival. Higher scores correlate with lower survivals (41).

[2] The *ASA score* is a six-category classification system for assessing the fitness of patients before surgery and anesthesia. Higher scores correlate with lower survivals.

Abbreviations: SD = standard deviation

3 Multidisciplinary evaluation of diagnostics

Assessment of the ^{18}F-FDG PET/CT and CECT scans within first-line follow-up was completed by a multidisciplinary expert panel. The panel, consisting of two nuclear physicians (LFG and DR) and two radiologists specialised in abdominal examinations (EPM and SSF) appointed tumour recurrences. For liver detail, all the scans were individually assessed by two members of the expert panel. The CECT scans were anonymized, shown in random order and analysed by the two experienced abdominal radiologists (SSF and EPM). The ^{18}F-FDG PET/CT scans were anonymized, shown in random order and analysed by the two experienced nuclear physicians (LFG and DR). All scans were checked for extrahepatic metastases by LFG, DR and alternately by EPM and SSF.

The radiologists were blinded for clinical data other than CEA level and previous scans. The nuclear physicians were blinded for clinical data other than CEA level and previous scans. However, the nuclear physicians were allowed to analyse the available CECT scan. In case of disagreement between two assessments of the same imaging modality, a second view of the same experts was realized and consensus was reached. Decisions by the multidisciplinary team meeting and/or follow-up imaging were considered golden standards since no standard biopsy was taken of suspected tumour lesions.

4 Disease states

Local/residual recurrence in the ablation or resection zone following LAT indicates the strength of the provided intervention. New intrahepatic and extrahepatic metastases following LAT indicate the degree of aggressiveness of colorectal liver disease. Early tumour detection ensures that we minimalize intrahepatic non-local and extrahepatic distant spread. Also, early detection ensures that tumours are relatively small and can be treated with minimally invasive interventions with minimal risks of adverse events.

Researcher SvM classified all unique and raw observations on both imaging modalities from the multidisciplinary evaluation of the first-line diagnostics to a set of mutually exclusive and collectively exhaustive disease states. Retrospectively used as a reference was the hospital's electronic patient database. Table 2 shows that the detection of CRLM was divided into four patient-level disease states: **(1)** no recurrence, **(2)** residual/local recurrence in the ablation or resection zone (=no new metastases but residual/local recurrence only demonstrating the strength of the provided intervention), **(3)** any intrahepatic metastases (≥1 new intrahepatic non-local metastases possibly supplemented with residual/local recurrence) and **(4)** any extrahepatic metastases (≥1 new extrahepatic distant metastases possibly supplemented with residual/local recurrence and/or new intrahepatic non-local metastases). Note that after classification, an expert panel consensus was reached if any disagreement occurred.

Table 2 Four disease states representing patient-level observations from the clinical trial

Observed in the clinical trial	**Disease state in the DES model**
No recurrence/clean follow-up	No recurrence **(1)**
Residual/local recurrence only	Residual/local recurrence **(2)**
Intrahepatic non-local metastases	Any intrahepatic metastases **(3)**
Residual/local recurrence *and* intrahepatic non-local metastases	Any intrahepatic metastases **(3)**
Distant metastases outside the liver	Any extrahepatic metastases **(4)**
Distant metastases outside the liver *and* residual/local recurrence	Any extrahepatic metastases **(4)**
Distant metastases outside the liver *and* intrahepatic non-local metastases	Any extrahepatic metastases **(4)**
Distant metastases outside the liver, residual/local recurrence *and* intrahepatic non-local metastases	Any extrahepatic metastases **(4)**

Estimating the consequences of false-negative and false-positive test results

Observations from the clinical trial made by the clinicians were essential input parameters for the DES model. In most cases, the observation was a true test result (i.e., the observation "no recurrence" would be a true negative test result and all other observations would be true positive test results) but some observations were false test results (i.e., observation "no recurrence" would be a false-negative test result and all other observations would be false-positive test results). Typically, each observation had a probability of being a true or false test result resulting in, for example: 'the number of patients with no recurrence = true negative test results + false-positive test results'. The concerned decision-making depends on the test result presented (Figure 1).

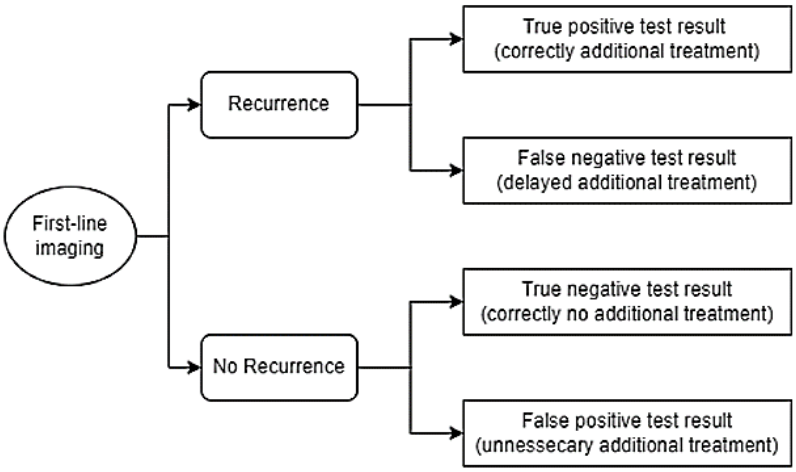

Figure 1 Clinical decision-making depends on true and false test results

Scenario analyses for false-negative test results

Scenario analyses were performed to assess the consequences of false-negative imaging resulting in a varying spread of CRLM recurrence. Proportions of occurring CRLM spread after a false-negative test result were restricted to the implementation of three 'what-if' scenarios (Table 3).

Table 3 Three what-if scenarios for estimating the consequences of false-negative test results

Scenario	Implementation
Worst case	All patients with a false-negative test result develop extrahepatic distant metastases
Base case	All patients with a false-negative test result have a probability of 1/3 to develop intrahepatic residual/local metastases, a probability of 1/3 to develop intrahepatic non-local metastases and a probability of 1/3 to develop extrahepatic distant metastases
Best case	All patients with a false-negative test result develop intrahepatic residual/local metastases only

5. Introduction to DES

DES provides a flexible framework that can be used to model a wide variety of health care problems (89–93). A patient-level and process-oriented DES model was developed to represent the complex dynamics of CRLM clinical practice and how clinical decision-making impacts cost-effectiveness

outcomes. Patient-level variation was reflected as it was relevant to address the differences in parameter values between subgroups (e.g., age dependence, LAT type received). DES is a modelling technique to which the challenges associated with discrete-time cycles do not apply. Events can occur at any time because the time to these events is modelled using smooth time-to-event distributions (86). Particularly in our scenarios in which few events would be observed per time cycle, the use of DES was preferable (e.g., compared to patient-level Markov modelling).

For all simulation analyses, *R* version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria; http://www.R-project.org/) was used, package *simmer* version 4.4.3 (94). For all cost-effectiveness analyses, *R* version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria; http://www.R-project.org/) was used, package *BCEA* version 2.4.1 (95).

6 Populating the DES model

All simulated patients entered the model immediately after diagnosis of CRLM. All patients received LAT preceded, if necessary, by induction therapy. There were two types of imaging modalities available for first-line diagnosis of CRLM recurrence: conventional CECT and ^{18}F-FDG PET/CT respectively. During higher-order surveillance of potential tumour recurrences, patients remained in a general follow-up state. After the diagnosis of recurrence, repeat LAT (rLAT) might be received. If rLAT would be futile, systemic therapy could be started and complemented with supportive end-of-life care.

A graphical representation of the DES model structure is provided in Figure 2. In upcoming paragraphs, the CRLM care pathway will be discussed chronologically and in more detail.

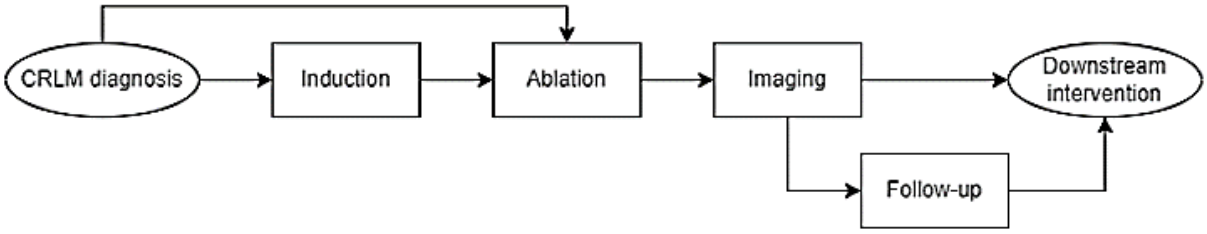

Figure 2 Graphical representation of the CRLM care pathway

Induction therapy and LAT strategies

All entered patients received LAT with curative intent (i.e., autonomous TA therapy or TA additional to metastasectomy). Percutaneous RFA and MWA procedures were performed under anaesthesia using ultrasound (US) and/or CT guidance for needle positioning and to assess the extent of the ablation zone. Open ablation was performed during surgery under general anaesthesia and US was used for needle positioning. The duration of ablation was determined by the discretion of the interventional radiologist and was documented in the treatment report. TA procedures were performed by three different experienced interventional radiologists (MB, CvR, AvE), assisted by technologists.

Clinical trials argue that LAT is costly and burdensome to the patient (7,9,10). The reported health and economic burden indicate the importance of strict identification of patients likely to benefit from receiving LAT. Also, LAT is associated with a substantial risk of treatment-related complications. Complication rates are based on reported estimates by the Dutch Hepato Biliary Audit (DHBA). Complications cause health and economic burden in terms of reduction in quality of life for the patients and additional healthcare costs. Therefore, we divided the costs related to LAT into fixed intervention costs per treatment type and variable additional hospitalization costs (i.e., the longer the hospitalization the more invasive the intervention and/or the more extensive the consequences of the complications).

In advance of LAT, some of these patients received induction therapy to shrink the tumours to be resectable during the intervention. Provided chemotherapy could have been CapOx, FOLFOX, FOLFIRI or FOLFIRINOX with or without Avastin, panitumumab or cetuximab. On average, induction therapy was provided for 16-20 weeks depending on the extent to which tumours shrink depicted on minimal two independent consecutive CECT scans. It was assumed that all patients receiving induction therapy responded to chemotherapy (i.e., the tumours shrank in terms of size and/or number). It was also assumed that no huge complications occurred during induction therapy such that an induction therapy regimen was never interrupted (e.g., in practice, with a dose reduction the patient could always complete the whole chemotherapy regimen).

First-line follow-up: adopting ^{18}F-FDG PET/CT imaging

Within 6 months following LAT, all patients were monitored for CRLM recurrence with a single ^{18}F-FDG PET/CT scan and conventional CECT imaging supplemented by serum CEA (in the remainder of this report referred to as **'the intervention group'**) or monitored with conventional CECT imaging supplemented by serum CEA (in the remainder of this report referred to as **'the comparator group'**). Patients with suspicious findings on these scans underwent an additional MRI Primovist or MRI abdomen. MRI was performed to further analyse suspected CRLM lesions before going for an additional consecutive treatment strategy. Patients, in whom MRI imaging was inconclusive for detecting or excluding CRLM, received a CT-guided puncture of suspicious liver tissue.

Within the intervention group, patients with elevated serum CEA (i.e., CEA level>3.8 µg/L) received a whole-body (skull to mid-thigh) ^{18}F-FDG PET/CT scan. Patients with normal CEA levels received ^{18}F-FDG PET/CT imaging of liver detail and thoracic-abdominal CECT. Combination ^{18}F-FDG PET/CT imaging (i.e., PET imaging and a simultaneous low-dose CT scan) was performed using the Philips Vereos Digital ^{18}F-FDG PET/CT system (Philips, Eindhoven, the Netherlands). FDG dose and acquisition time was calculated based on body weight and height.

After first-line follow-up, a patient was transferred to a downstream intervention strategy if CRLM recurrence occurred (=disease states **2-4**). All patients with no recurrence (=disease state **1**) were transferred to higher-order follow-up lines.

Multiple lines of higher-order follow-up diagnostics: a general follow-up phase

By hospital's protocol and Dutch guidelines, the patients with no recurrence (=disease state **1**) after a completed first-line follow-up workup received a higher-order follow-up schedule consisting of 3-4 monthly serum CEA and CECT scans during years 1-2 and 6 monthly serum CEA and CECT scans during years 3-5. Patients with suspicious findings on these scans underwent an additional MRI Primovist or MRI abdomen and/or ^{18}F-FDG PET/CT scan. These scans were performed to further analyse suspected CRLM lesions before going for an additional downstream treatment strategy. Patients, in whom MRI and/or PET/CT imaging was inconclusive for detecting or excluding CRLM, received a CT-guided puncture of suspicious liver tissue.

Higher-order follow-up data was collected with literature evidence on tumour progression rates and progression-free survival (PFS) probabilities (Table 4). Patients with initially no recurrence were at risk of developing recurrence over at most twelve years. In our model, developing recurrence would cause patients to transfer to a downstream intervention strategy. Otherwise, if the patient was disease-free over twelve years, it was concluded that the patient was thoroughly clean, no CLRM will be developed in the future and the patient simply followed the Dutch life tables by age-dependent overall survival times. We have retrieved and averaged survival times from these Dutch life tables for the period 2016-2020. (96)

Table 4 PFS probabilities after a completed higher-order follow-up schedule over twelve years

Recurrence developed	**Value** [1]	**Uncertainty** [2,3]	**Source**
No recurrence	17% (24/144)	Dirichlet (*alpha*=[24,21,50,49])	(29,60)
Residual/local recurrence	14% (21/144)	Dirichlet (*alpha*=[24,21,50,49])	(29,60)
Any intrahepatic metastases	35% (50/144)	Dirichlet (*alpha*=[24,21,50,49])	(29,60)
Any extrahepatic metastases	34% (49/144)	Dirichlet (*alpha*=[24,21,50,49])	(29,60)

[1] Base-case values (percentage, absolute number of total observations), i.e., no parameter uncertainty incorporated

[2] Dirichlet distribution is a natural choice because a direct match with literature evidence

[3] Distribution parameters are presented as *alpha* which is one set of parameters for all observations

In each disease state, a probability of death due to other causes was derived from available Dutch demographic data (96). This all-cause probability of death was assigned to each patient based on the time spent in follow-up and the age of the patient. We retrieved and averaged these general Dutch age-dependent survival times for the period 2016-2020. These data can be found in *Appendix A.*

Downstream treatment strategies

All patients with diagnosed recurrent CRLM (=disease states **2-4**) were transferred to downstream intervention strategies. Based on the spread of the detected recurrence, treatment history (e.g., recurrent disease) and general patient characteristics (e.g., age, clinical condition), the downstream intervention strategy was determined during multidisciplinary review board meetings. Table 5 visualizes the resulting dominant and recessive downstream treatment strategies per disease state.

Table 5 Representing the dominant and recessive downstream treatment strategies per type of CRLM recurrence detected

Recurrence demonstrated on received diagnostics	**Repeat LAT [1]**	**Systemic therapy [1]**	**Wait-and-see management [1]**	**Uncertainty [2,3]**
No recurrence	-	-	100% (32/32)	Beta (α=32, β=0)
Residual/local recurrence	64% (7/11)	36% (4/11)	-	Beta (α=7, β=4)
Any intrahepatic metastases	78% (14/18)	22% (4/18)	-	Beta (α=14, β=4)
Any extrahepatic metastases	24% (7/29)	76% (22/29)	-	Beta (α=7, β=22)

[1] Base-case values (percentage, absolute number of total observations), i.e., no parameter uncertainty incorporated

[2] Beta distribution is a natural choice because a direct match with clinical evidence

[3] Distribution parameters are presented as α- and β-parameters for beta distributions

In a favourable case, the spread of CRLM recurrence was limited and rLAT could be provided. Downstream rLAT procedures may consist of repeat TA therapy, partial hepatectomy or stereotactic body radiation therapy. If rLAT would no longer be effective, systemic therapy could be provided. Systemic therapy often consisted of multiple consecutive chemotherapy regimens. The most common received chemotherapy was Capecitabine/Oxaliplatin with Avastin or Capecitabine/Avastin with sequentially Oxaliplatin. An alternative was a regimen with Irinotecan, monotherapy or FOLFIRI, panitumumab or cetuximab, or optionally LONSURF.

In each downstream intervention, a probability of death due to other causes was derived from available Dutch demographic data (96). This all-cause probability of death was assigned to each patient based on the time spent in downstream treatment strategies and the age of the patient. We retrieved and averaged these general Dutch age-dependent survival times for the period 2016-2020 (*Appendix A*).

Moreover, after a longer period of downstream intervention strategies, it was assumed that each patient did not survive. The sum of the disease-related and the all-cause probability of dying was 100%. So that it was assumed that all patients arriving in downstream intervention strategies received end-of-life supportive care sequentially after rLAT or systemic therapy.

7 Survival analysis

Survival data per downstream intervention strategy was collected from literature (12,19,29,60). The results of literature evidence (e.g., randomized controlled trials) on time-to-event outcomes, usually reported, were median time-to-events and Cox Hazard ratios. These did not constitute the sufficient statistics required for cost-effectiveness analysis. Also, the selection of the type of time-to-event distributions could have a major impact on outcomes when extrapolating beyond the time horizon supported by the data. Given our limited follow-up time in the single-centre LUMC cohort data, it may be expected that the uncertainty in estimates of hazards increased when extrapolating the further into the future. Consequently, the magnitude of uncertainty in estimates of a lifetime mean survival and cost-effectiveness increased when extrapolating the further into the future. (97–102)

The Kaplan-Meier (KM) estimator is a non-parametric statistic used to estimate the survival function from lifetime data. When no truncation or censoring occurs, the KM curve is the complement of the empirical distribution function. However, specified individual patient-level time-to-event data and the status at the last observation are required to generate such a KM estimator. Consequently, for this type of survival analysis, there is a need for an empirical baseline hazard function. To enhance the quality of secondary data analyses, we applied a method which derives from published KM survival curves a close approximation to the original individual patient-level time-to-event data from which they were generated. (103,104) This algorithm of *Guyot et al*. maps from digitized curves back to KM data by finding numerical solutions to the inverted KM equations using the information on the number of events and the numbers at risk. We checked and justified which distribution was best by visual checks (e.g., Q-Q plots, P-P plots, histograms and density plots) and statistical checks (e.g., AIC, BIC). The resulting KM data are presented in *Appendix B.*

8 Model parameters

The simulation model was structured using a wide range of input parameter values. Model parameter values that were used in the base case, mostly, were a direct match with (binomial) available data from the single-centre clinical trial complemented by retrospectively searching the hospital's electronic patient database. Missing hospital data were the diagnostic performance of conventional CECT, respectively, ^{18}F-FDG PET/CT imaging, survival times and utility rates. These model input parameters, and costs of care as usual, were derived from (aggregated) literature evidence and from Statistics Netherlands adhered to the Dutch national guidelines. The final leftovers evidence gaps (i.e., parameters for which no information was found or that varied highly among literature) were addressed and valued by an expert panel consisting of two nuclear physicians (LFG and DR), two radiologists specialised in abdominal examinations (EPM and SSF), one interventional radiologist (MB), one surgeon specialized in abdominal oncology (SDM) and a Full Professor health technology and services research (HK).

A description of the evidence, that was provided in the trial data, is presented in Table 6.

First-line follow-up: adopting ^{18}F-FDG PET/CT imaging

The model included a complete first-line follow-up of the entire study population (i.e., all test results from the clinical trial were noted and, for each patient, it was concluded if CRLM recurrence occurred). A description of the trial evidence on tumour recurrence developed during first-line follow-up is listed in Tables 7-8. Table 7 shows the recurrence detected by the comparator group. Table 8 covers the same patients, but the recurrence detected is reclassified for the intervention group.

Table 6 Evidence for DES modelling gathered as clinical trial data

Name	Data type	Information	Value [1]	Uncertainty [2]	Distr.
RID	Integer number	Patient's unique research identifier	-	-	-
Age	Number of years	Patient's age	65	shape=6.54, scale=69.39	Weibull
Colon	Binary (1=colon, 0=rectal)	Patient's primary CRC site	67% (30/45)	α=30, β=15	Beta
CEA level	Binary (1=stable, 0=decreased)	Patient's CEAs predict recurrence	69% (22/32) [3]	α=22, β=10	Beta
Fong score	Binary (1=high, 0=low)	Patient's clinical risk score	44% (14/32) [3]	α=14, β=18	Beta
ASA score	Binary (1=high, 0=low)	Patient's clinical condition	33% (15/45)	α=15, β=30	Beta
Recurrent	Binary (1=recurrent, 0=first)	First diagnosis or recurrent disease	51% (23/45)	α=23, β=22	Beta
Induction therapy	Binary (1=yes, 0=no)	Patient received induction therapy	29% (13/45)	α=13, β=32	Beta
Response to induction therapy	Binary (1=yes, 0=no)	Patient's response induction therapy	100%	Fixed [4]	-
Complications in induction therapy	Binary (1=yes, 0=no)	Nominative complications	0%	Fixed [4]	-
LAT procedure type	Binary (1=autonomous TA, 0=TA additional surgery)	Patient received autonomous TA	76% (34/45)	α=34, β=11	Beta
Complications reported by the DHBA [5]	Binary (1=yes, 0=no)	Complications *if autonomous TA*	21% (7/34)	α=7, β=27	Beta
		if TA additional to metastasectomy	28% (3/11)	α=3, β=8	Beta
MRI provided per follow-up line	Binary (1=yes, 0=no)	Additional MRI for surveillance	24% (11/45)	α=11, β=34	Beta
Biopsy provided per follow-up line	Binary (1=yes, 0=no)	Additional biopsy for surveillance	9% (4/45)	α=4, β=41	Beta
^{18}F-FDG PET/CT provided for each higher-order FU	Binary (1=yes, 0=no)	^{18}F-FDG PET/CT adopted in higher order surveillance	24% (11/45)	α=11, β=34	Beta

[1] Base-case values (percentage, absolute number of total observations), i.e., no parameter uncertainty incorporated; [2] Distribution parameters are presented as α- and β-parameters for beta distributions; [3] The CEA levels and Fong scores are not always measured/reported (i.e., 13/45 missing values); [4] Estimated after elicitation from the previously mentioned expert panel; [5] Having complications is defined as "complication" by the DHBA resulting in a prolonged hospital stay.

Table 7 Comparator group: CRLM recurrence detected in first-line follow-up

Detection of CRLM recurrence	Value [1]	Uncertainty [2,3]
No recurrence/clean first-line follow-up	38% (17/45)	Dirichlet (*alpha*=[17,4,9,15])
Residual/local recurrence detected	9% (4/45)	Dirichlet (*alpha*=[17,4,9,15])
Any intrahepatic non-local metastases detected	20% (9/45)	Dirichlet (*alpha*=[17,4,9,15])
Any extrahepatic distant metastases detected	33% (15/45)	Dirichlet (*alpha*=[17,4,9,15])

[1] Base-case values (percentage, absolute number of total observations), i.e., no parameter uncertainty incorporated

[2] Dirichlet distribution is a natural choice because a direct match with literature evidence

[3] Distribution parameters are presented as *alpha* which is one set of parameters for all observations

Table 8 Intervention group: reclassified CRLM recurrence detected in first-line follow-up

Reclassified detection of CRLM recurrence	Value [1]	Uncertainty [2,3,4]
If no recurrence was detected by the comparator group		
No recurrence/clean first-line follow-up	88% (15/17)	Dirichlet (*alpha*=[15,ε,1,1])
Residual/local recurrence detected	0% (0/17)	Dirichlet (*alpha*=[15,ε,1,1])
Any intrahepatic non-local metastases detected	6% (1/17)	Dirichlet (*alpha*=[15,ε,1,1])
Any extrahepatic distant metastases detected	6% (1/17)	Dirichlet (*alpha*=[15,ε,1,1])
If residual/local recurrence was detected by the comparator group		
No recurrence/clean first-line follow-up	0% (0/4)	Dirichlet (*alpha*=[ε,4,ε,ε])
Residual/local recurrence detected	100% (4/4)	Dirichlet (*alpha*=[ε,4,ε,ε])
Any intrahepatic non-local metastases detected	0% (0/4)	Dirichlet (*alpha*=[ε,4,ε,ε])
Any extrahepatic distant metastases detected	0% (0/4)	Dirichlet (*alpha*=[ε,4,ε,ε])
If intrahepatic non-local recurrence was detected by the comparator group		
No recurrence/clean first-line follow-up	0% (0/9)	Dirichlet (*alpha*=[ε,ε,8,1])
Residual/local recurrence detected	0% (0/9)	Dirichlet (*alpha*=[ε,ε,8,1])
Any intrahepatic non-local metastases detected	89% (8/9)	Dirichlet (*alpha*=[ε,ε,8,1])
Any extrahepatic distant metastases detected	11% (1/9)	Dirichlet (*alpha*=[ε,ε,8,1])
If extrahepatic distant recurrence was detected by the comparator group		
No recurrence/clean first-line follow-up	0% (0/15)	Dirichlet (*alpha*=[ε,3,ε,12])
Residual/local recurrence detected	20% (3/15)	Dirichlet (*alpha*=[ε,3,ε,12])
Any intrahepatic non-local metastases detected	0% (0/15)	Dirichlet (*alpha*=[ε,3,ε,12])
Any extrahepatic distant metastases detected	80% (12/15)	Dirichlet (*alpha*=[ε,3,ε,12])

[1] Base-case values (percentage, absolute number of total observations), i.e., no parameter uncertainty incorporated; [2] Dirichlet distribution is a natural choice because a direct match with literature evidence; [3] Distribution parameters are presented as *alpha* which is one set of parameters for all observations; [4] The value ε represents a small and randomly selected number close to zero, and, restricted for $0<\varepsilon<0.5$.

Patient-level diagnostic performance of imaging modalities

Literature-based evidence on the evaluation of the diagnostic accuracy of ^{18}F-FDG PET/CT and CECT for the detection of disease progression following TA therapy was included in our model (15,105). *Samim et al.* reported on nine studies defining the diagnostic accuracy of ^{18}F-FDG PET/CT imaging

following TA and seven studies defining the accuracy of conventional CECT imaging following TA. *Samim et al.* performed subgroup analysis for different time intervals between TA therapy and the follow-up imaging for evaluation of CRLM recurrence **(1)**, lesion and patient-based analysis **(2)**, and including/excluding patients with non-intrahepatic mCRC **(3)**. We selected the accuracy values of our interest, which were patient-level sensitivity and specificity values with a time interval of at most 6 months between TA therapy and follow-up imaging, and, including patients with potentially detected extrahepatic metastases (i.e., not oligometastatic liver disease only). Table 9 shows these sensitivity and specificity values. Note that the evidence from Table 9 represents asymmetric observational data.

Table 9 Patient-level diagnostic performance of conventional CECT and when adopting ^{18}F-FDG PET/CT respectively

	Sensitivity (95%-CI)	**Specificity (95%-CI)**
Conventional CECT workup	81.8% (71.6-89.0)	98.2% (91.5-99.6)
^{18}F-FDG PET/CT workup	91.6% (83.7-95.8)	97.2% (90.8-99.2)

The multivariate normal distribution was used to approximate any set of correlated random sensitivity and specificity values. The approximated normally distributed collection of correlated sensitivity and specificity values were projected on Beta distributions so that we avoided unrealistic extreme sensitivity and specificity rates (which were negative values and values≥1 respectively). Table 10 shows the mean, median, 25th percentile and 75th percentile of the approximated and beta-distributed sensitivity and specificity values used for simulation modelling.

Table 10 Summary statistics of the sensitivity and specificity values used for simulation

	Simulated sensitivity		**Simulated specificity**	
Conventional CECT workup	mean	85%	mean	98%
	25th percentile	*83%*	*25th percentile*	*98%*
	Median	*99%*	*Median*	*99%*
	75th percentile	*99%*	*75th percentile*	*99%*
^{18}F-FDG PET/CT workup	mean	91%	mean	97%
	25th percentile	*90%*	*25th percentile*	*96%*
	Median	*99%*	*Median*	*99%*
	75th percentile	*99%*	*75th percentile*	*99%*

We assumed that the correlation between the sensitivity and specificity is the same for the intervention group and comparator group. Furthermore, we assumed that the sensitivity of the intervention group at least equals the sensitivity of the comparator group (correlation=1). That is because if we adopt an extra imaging modality to routine first-line follow-up, we are always going to detect more lesions suspected of recurrence. As a result, the number of false-negative test results in the intervention group will never exceed the number of false-negative test results in the comparator group.

Costs

Costs were determined from a healthcare perspective including the costs related to treatments and diagnostics (106). Treatment costs were approximated utilizing the hospital's specific pricing included in the Dutch system of Diagnosis-Treatment Combinations (DBC). Costs for complications during treatment were expressed in prolonged hospitalization and/or additional interventional procedures received. Independent imaging costs were derived from the hospital's reference prices with an effective date of July first, 2021. Furthermore, costs for end-of-life supportive care were assigned to patients who received any downstream patient management.

An overview of all included health-related costs is given in Table 11. Note that pricing agreements may differ per hospital. Following the Dutch guidelines for health economic evaluations, annual discount rates of 4.0% were applied to all future costs (107–109).

Table 11 Cost parameter values approximated utilizing the hospital's specific DBC pricing

Cost parameter	DBC pricing [1]	Source
CECT thorax-abdomen	€234.98	(110)
^{18}F-FDG PET/CT (whole body)	€1,298.82	(110)
MRI Primovist/MRI abdomen	€384.82	(110)
CT-guided puncture liver	€935.32	(110)
Induction therapy (whole regimen)	€2,224.93 [2]	(110)
TA/low invasive metastasectomy	€11,547.05 [2]	(110)
Short hospital stay (<6 days)	€3,221.56 [2]	(110)
Prolonged hospital stay (<29 days)	€9,038.56 [2]	(110)
Extended hospital stay (≥29 days)	€39,391.09 [2]	(110)
rLAT procedure (averaged)	€20,318.41 [2]	(110)
Systemic therapy regimen (averaged)	€10,672.41 [2]	(110)
Supportive care (end-of-life care)	€6,088.78 [2]	(110)

[1] Reflected hospital's cost values for groups of patients (that is the average costs)

[2] Including costs of additional medication, complications solving, pain management and examinations

Health effects

Health-related utilities were derived from literature (111,112). The applied literature-based evidence on utilities was calculated from EQ-5D index scores and can be found in *Appendix C*. These utilities represent the valuation of health-related quality of life (HR-QoL) over time on a scale from zero (death) to one (perfect health). All evidence presented in *Appendix C* is averaged in Table 12. From the mean utility values for each disease state, it was clear that the disease-free group (=disease state **1**) experienced a better quality of life than the group with recurrent CRLM (=disease states **2-4**). (112)

Disutilities were subtracted if treatment-related complications appeared. The disutility in case of complications was, on average, valued as 10% of perfect health, which was an assumption based on aggregated literature evidence and elicitation from the previously mentioned expert panel. (111)

Table 12 Mean health-related quality of life values for each disease state (112)

Health state	Mean value (SD)	No. observations
Death	0 (0)	349
No recurrence (=disease state **1**)	0.78 (0.23)	891
Recurrence (=disease states **2-4**)	0.74 (0.25)	450
▪ LAT received	0.82 (0.17)	205
▪ Systemic therapy received	0.68 (0.28)	245
Non-curative care	0.67 (0.31)	162

Finally, QALYs were calculated by the discounted sum of utilities over the lifelong evaluation period. Following the Dutch guidelines for health economic evaluations, annual discount rates of 1.5% were applied to all future health outcomes (107–109).

9 Handling competing risks

Different strategies for implementing competing risks in DES models are available. We decided to select our event first and the time-to-event second, because of its easy implementation. This approach strictly divided the data according to different event types and, consequently, was sensitive to low event rates which might affect the performance of our model negatively (113).

To select events, we draw random numbers from the continuous uniform distribution with a specified range of (0,1). The way an event was determined based on a random draw and competing risks, is visualised in Figure 3. In this figure, for example, three events could occur with probability p1, p2, p3 and p1+p2+p3=1 respectively. This principle was used for all events and competing risks in our model.

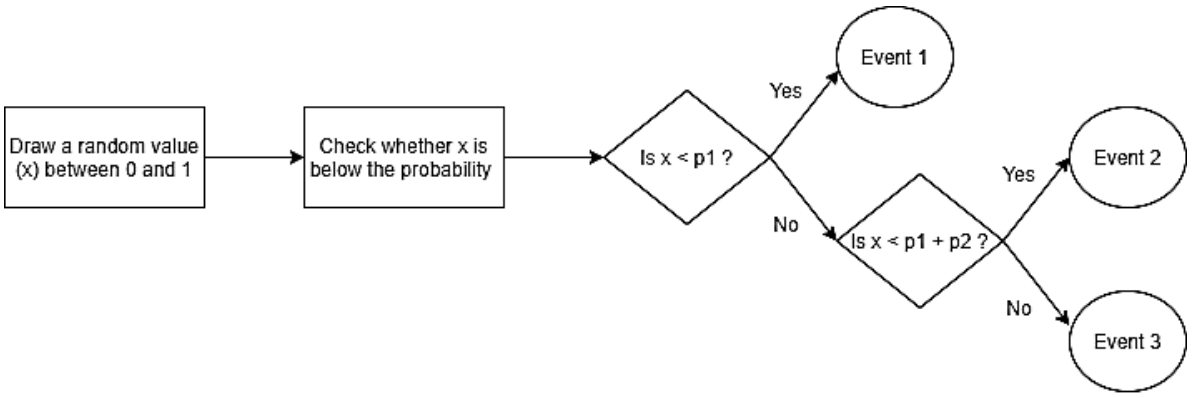

Figure 3 Comparing strategies for modelling competing risks: event first, time-to-event second

Reflecting variations in time-to-event data

A three-step approach was applied for using parametric probability distributions, which consisted of fitting the distributions **(1)**, checking the distribution fit **(2)** and drawing values from the distributions in the simulation model **(3)**. Checking the distribution fit was ensured in two ways: a visual check by histogram density function overplots and a statistical check by the Akaike information (AIC) and

Bayesian information criteria (BIC). In some cases, variables were bound by natural limits. If so, truncated distributions were used. A summary of the three-step approach is given in Table 13.

Table 13 Three-step approach for using parametric probability distributions

Three-step approach	**Continuous non-censored data or fitting distributions based on summary statistics**
Fitting the distribution	(1) Inspect the data (2) Fit multiple distributions
Checking the distribution fit	(1) Provide a goodness-of-fit test (2) Provide density functions histogram (3) Determine the statistical error AIC/BIC
Propagate uncertainty	(1) Trunk for extreme values (test for n=10,000) (2) Draw a random value from the corresponding distribution (3) Evaluate the model (run ≥ 1000 times)

Typically, time-to-event distributions are survival distributions defined by two parameters (e.g., shape and scale). By subsetting the data based on the event of interest, time-to-event data was derived from the hospital's clinical trial and complemented by a literature search. An overview of reflected variations in the time-to-event data for the corresponding model parameters is given in Table 14.

Table 14 Reflecting variations in time-to-event data

Time-to-event parameter	**Value [1]**	**Uncertainty**	**Distribution**	**Source**
Response to induction therapy	195	meanlog=5.24, sdlog=0.29	Lognorm [5]	[7]
Hospitalization following LAT				
if autonomous TA therapy	1	meanlog=0.060, sdlog=0.30	Lognorm	[7]
if complications	11	shape=0.83, scale=9.22	Weibull [5]	[7]
if TA and metastasectomy	6	meanlog=1.67, sdlog=0.34	Lognorm	[7]
if complications	17	shape=0.83, scale=9.22	Weibull [5]	[7]
First-line follow-up by the comparator group [2]	118	mean=117.71, sd=28.19	Normal [5]	[7]
First-line follow-up by the intervention group [2]	118	mean=118.33, sd=26.66	Normal [5]	[7]
Progression-free survival (PFS) [3]				
if local recurrence	284	meanlog=1.98, sdlog=0.73	Lognorm [5,6]	(29,60)
if non-local/distant recurrence	324	meanlog=2.00, sdlog=0.87	Lognorm [5,6]	(29,60)
Overall survival (OS) [4]				
if rLAT received	1402	shape=2.93, scale=37.81	Log-logistic [6]	(29,60)
if systemic therapy received	1266	shape=2.51, rate=0.060	Gamma [6]	(12,19)

[1] Base-case values *in days*, i.e., no parameter uncertainty incorporated; [2] Time-to-event data for all test results (no, local, non-local/distant recurrence detected respectively); [3] PFS probabilities of follow-up consisting of multiple lines of higher-order diagnostics scheduled; [4] Overall survival times during downstream intervention strategies consisting of repeat local and systemic interventions respectively; [5] Truncated distribution for lower and/or upper extreme (and negative) values following the hospital's protocol; [6] The selection of time-to-event distributions can have a major impact on outcomes when extrapolating; [7] Direct match with clinical trial data complemented by the hospital's electronic patient database.

10 Cost-effectiveness analysis

Incremental cost-effectiveness ratios (ICERs) were calculated by dividing the estimated mean difference in costs by the mean difference in quality-adjusted life years (QALYs). Health outcomes were also expressed in life years gained (LYGs) due to markedly variations in reported HR-QoLs in literature. Additionally, as a function of the willingness to pay (WTP) for a QALY, the cost-effectiveness acceptability curve (CEAC) was used to graph the probability that the intervention group is cost-effective compared to the comparator group. (114–117) In the Netherlands, WTP thresholds of €20,000 and €50,000 per QALY are recommended by the Dutch Council for Public Health and Health Care for conditions with an intermediate disease burden (108).

Note that heterogeneous individuals moved through our DES model. Recording individual patient outcomes may be interesting, however, the clinical decision-making to be supported will be on cohort level (i.e. for the entire (sub)group of patients). Therefore, we assumed that individuals and society wish to maximize their health outcomes and aim to achieve the greatest benefit for the greatest number. (116) Moreover, in literature, it is argued that rules of inference are arbitrary and entirely irrelevant to the decisions which clinical and economic evaluations claim to inform. Decisions should be based only on the mean net benefits irrespective of whether differences are statistically significant or fall outside a Bayesian range of equivalence. (118)

11 Probabilistic analyses

Uncertainty refers to the fact that we cannot know with absolute certainty what the expected effects and costs of an intervention are. So, there will always be a chance that the wrong adoption decision is made resulting in potentially harming patients due to the consequences of the uncertainty. (89,119)

The consequences of such uncertainty in input parameter values on model outcomes were assessed through probabilistic analysis (PA). PA was performed to reflect parameter uncertainty in the accumulated evidence used for the simulation analyses. The provided PAs were based on Monte Carlo simulations with 10,000 samples (=patients), which was a sufficiently large number to remove the impact of patient-level variation in the cost-effectiveness outcomes (see *Appendix D*).

Five independent PAs were provided: the base case and varying scenarios (Table 15). Two scenarios reflected different levels of tumour spread after a false-negative test result and before receiving (delayed) treatment. Another two scenarios reflected prolonged survival times in case of favourable downstream rLAT received.

Table 15 PA preparation runs, the base case and the four scenarios listed

PA scenario	Implementation
PA preparation runs	We determined the number of samples (=patients) to simulate, which should be a sufficiently large number to remove the impact of patient-level variation (i.e. stochastic uncertainty) in cost-effectiveness outcomes. In this analysis, we did not incorporate parameter uncertainty (see *Appendix D* for results).
PA – the base case	PA base case reflected that all patients with a false-negative test result have a probability of 1/3 to develop intrahepatic residual/local metastases, a probability of 1/3 to develop intrahepatic non-local metastases and a probability of 1/3 to develop extrahepatic distant metastases before (delayed) treatment is received.
Scenario A1 – the worst case for false-negative test results	Scenario A1 reflected that all patients with a false-negative test result develop extrahepatic distant metastases before (delayed) treatment is received.
Scenario A2 – the best case for false-negative test results	Scenario A2 reflected that all patients with a false-negative test result develop intrahepatic residual/local metastases only (i.e., no further non-local/distant tumour spread) before (delayed) treatment is received.
Scenario B1 – prolonged survival if favourable rLAT was provided	In scenario B1 we assigned a prolonged survival time of 12 months if downstream rLAT was provided. In this way, we ensured that receiving downstream rLAT was more favourable than receiving downstream systemic therapy. The value of 12 months was based on literature evidence of the medians of overall survival times regarding both strategies (12,19,29,60).
Scenario B2 – alternative for applied survival distribution if rLAT was provided	Scenario B2 reflected the consequences of applying another time-to-event distribution for estimating overall survival times following downstream rLAT. In this scenario, we did not apply the log-logistic distribution as a parametric model for mortality rates, which initially increase and decrease later following LAT and CRLM restaging, but we used gamma-distributed time-to-event data defined by the parameters *shape=3.20* and *rate=0.074*.

12 Value of additional information

Uncertainty analysis can serve two main purposes: assess confidence in a chosen course of action **(1)** and ascertain the value of collecting additional information to better inform the decision **(2)**. In literature, it is argued that the evaluation of point estimates and uncertainty in parameters is part of a single process and explores the link between parameter uncertainty through decision uncertainty and the relationship to value of information analysis (89). Value of information (VOI) analysis gives insights into the amount a decision-maker should be willing to pay for more information before making a clinical decision to reduce the risk and consequences of making a wrong decision.

After the PA was performed to characterize the decision uncertainty, we established the value of additional information. We determined the VOI in terms of the expected value of perfect information

(EVPI). The EVPI for an individual patient is the difference between the expected benefits of making a decision with perfect and current information available. We used the EVPI as a first hurdle in recommending further research because the EVPI is widely accepted as a criterion for ruling out research that would not be worthwhile. (120–122) In terms of research prioritization decisions, the EVPI is argued to be the most appropriate presentational technique, alongside CEACs, for representing decision uncertainty from the PA (89).

The EVPI was determined by three parameters: the cost-effectiveness estimates given the information available **(1)**, the uncertainty surrounding these cost-effectiveness estimates **(2)** and the opportunity loss determined by the WTP threshold **(3)**. The EVPI could, without additional modelling, directly be calculated from the simulated results (i.e., QALYs combined with costs to inform decisions). We interpreted the EVPI as an upper bound on the returns from further research. If the population EVPI is bigger than the cost of additional research, then it is potentially cost-effective to do further research. When multiplying the EVPI for an individual patient by the expected population of patients who will benefit from the information, the maximum VOI derived from future research can be quantified. However, this required an estimate of the time over which the information would be beneficial **(1)**, the number of patients concerned **(2)** and the discounting of present values **(3)**. (122) Unfortunately, this information was not available. Therefore, we limited our VOI analysis by calculating the EVPI for an individual patient.

Results

The developed DES study consisted of multiple PAs. First, we provide PA cost-effectiveness outcomes for the base case. Afterwards, we explore PA cost-effectiveness estimates through scenario analyses.

PA outcomes: the base case

The mean number of events associated with the intervention and comparator group respectively, when simulating 10,000 runs of 10,000 patients with CRLM, is presented in Table 16. To point out, the mean number of residual/local recurrence observed in first-line follow-up in the intervention group compared to the comparator group was nearly doubled (1530 versus 888 cases). Further, the mean number of false-negative test results in first-line follow-up in the intervention group compared to the comparator group was nearly halved (737 versus 1374 cases). Lastly, on average, we observed a small increase in the number of consecutive LAT procedures provided in the case of the intervention group (4859 versus 4755 cases) with the consequence that we could waive (or postpone) providing less favourable complex systemic therapy in a small number of patients (for at least 1% of the total simulated population).

Table 16 Mean number of events per strategy when simulating 10,000 runs of 10,000 patients

	Intervention group	**Comparator group**
Description	*Number of events [1] (%)*	*Number of events [1] (%)*
No recurrence observed in first-line follow-up (FU1)	3382 (34)	3783 (38)
Residual/local recurrence observed in FU1	**1530 (15)**	**888 (9)**
Any intrahepatic non-local recurrence observed in FU1	2003 (20)	1997 (20)
Any extrahepatic distant recurrence observed in FU1	3085 (31)	3332 (33)
True test result in FU1	9197 (92)	8582 (86)
False-negative test result in FU1	**737 (7)**	**1374 (14)**
False-positive test result in FU1	66 (1)	44 (0)
No downstream treatment received [2]	*452 (4)*	*409 (4)*
Downstream rLAT received	**4859 (49)**	**4755 (48)**
Downstream systemic therapy received	4689 (47)	4836 (48)

[1] The mean number of events when simulating 10,000 runs of 10,000 individual patients with CRLM; [2] It should not be allowed that the intervention and comparator group influence the mean number of patients developing no recurrence resulting in an undesirable varying number of patients receiving no downstream treatment (elaborated in section 'Discussion').

The intervention group was estimated to provide 3.38 QALYs compared to 3.42 QALYs provided in the case of the comparator group over a lifetime horizon. Moreover, the intervention group was estimated to provide 5.08 LYGs compared to 4.98 LYGs in the case of the comparator group.

Costs associated with LAT received and consecutive diagnosis of recurrent CRLM were estimated to be higher for the intervention group compared to the comparator group (€20,419.10 versus €18,870.90). However, when accounting for the aforementioned costs and including the estimated costs of downstream patient management, the care pathway costs estimated appeared to be lower for the intervention group compared to the comparator group (€39,254.01 versus €40,681.57).

Finally, the summarised lifetime health and economic consequences in the case of the intervention group compared to the comparator group resulted in an ICER of €17,850.12 saved per QALY lost which was a non-dominant ratio, respectively, €14,668.67 saved per LYG which was a dominant ratio (Table 17).

Table 17 Costs, QALYs and LYGs per strategy resulting in the base case ICER per QALY and LYG respectively

Strategy	**Costs**	**QALYs**	**LYGs**	**ICER** *(€/QALY)*	**ICER** *(€/LYG)*
Intervention group	€39,254.01	3.38	5.08	intervention group non-dominant **€17,850.12 saved per QALY lost**	intervention group dominant **€14,668.67 saved per LYG**
Comparator group	€40,681.57	3.42	4.98		

The joint distribution in Figure 4 presents the incremental costs (in Euros) and health benefits (in QALYs) of the intervention group compared to the comparator group. Incremental cost-effectiveness point estimates above the horizontal can be considered cost increasing and point estimates to the right

of the vertical can be considered clinically beneficial. Point estimates below the diagonal represent simulations in which the intervention group was a cost-effective alternative compared to the comparator group given the Dutch WTP threshold of €20,000/QALY. The ellipsoid shape of the joint distribution indicates a clear correlation between incremental costs and incremental effects.

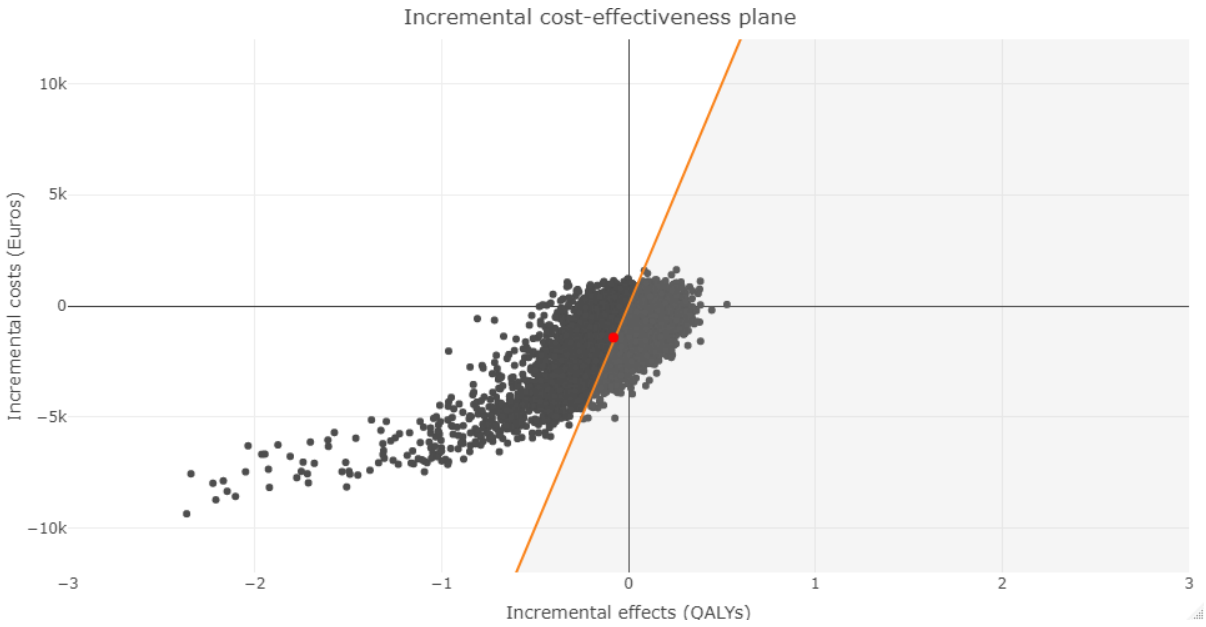

Figure 4 The incremental cost-effectiveness plane presents the ICER obtained through PA. Point estimates below the diagonal represent simulations in which the intervention group was cost-effective at a WTP threshold of €20,000/QALY

We transformed the results of the cost-effectiveness plane to the CEAC. Figure 5 shows the CEAC as a function of the WTP for a QALY. It shows that the intervention group was cost-effective at WTP thresholds of €20,000/QALY and €50,000/QALY in 54.0% and 43.6% of the simulations respectively. Accordingly, the comparator group was estimated to be the preferred strategy over the intervention group at the aforementioned WTP thresholds in 46.0% and 56.4% of the simulations respectively.

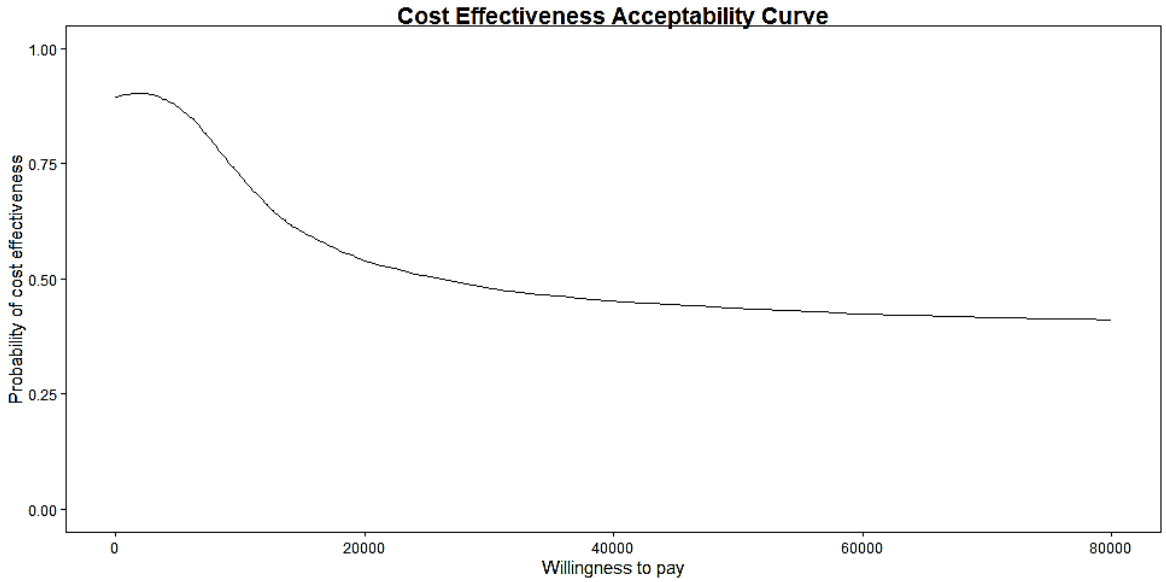

Figure 5 The CEAC assesses the probability that the intervention group was the cost-effective strategy as a function of the WTP threshold (for the base case PA)

Figure 6 shows the EVPI at different WTP thresholds. The EVPI increases as the threshold increases. The reason for this is that the EVPI increases with increasing uncertainty and with increasing consequences of making a (wrong) decision. Figure 6 demonstrates that the EVPI shows a nearly exponential rise, after which the curve flattens. Up to the point where the threshold value equals the value of the ICER (=where the dashed lines meet), the EVPI is nearly exponentially increasing because the uncertainty surrounding the adoption decision is increasing (=error probability is increasing), as is the value applied to the consequences of making an incorrect decision. After this point, the uncertainty in the adoption decision is flattening while the consequences associated with making an (in)correct decision continue to rise. It demonstrates that the falling, however nonnegligible, error probability is being outweighed by the costs of making an incorrect decision.

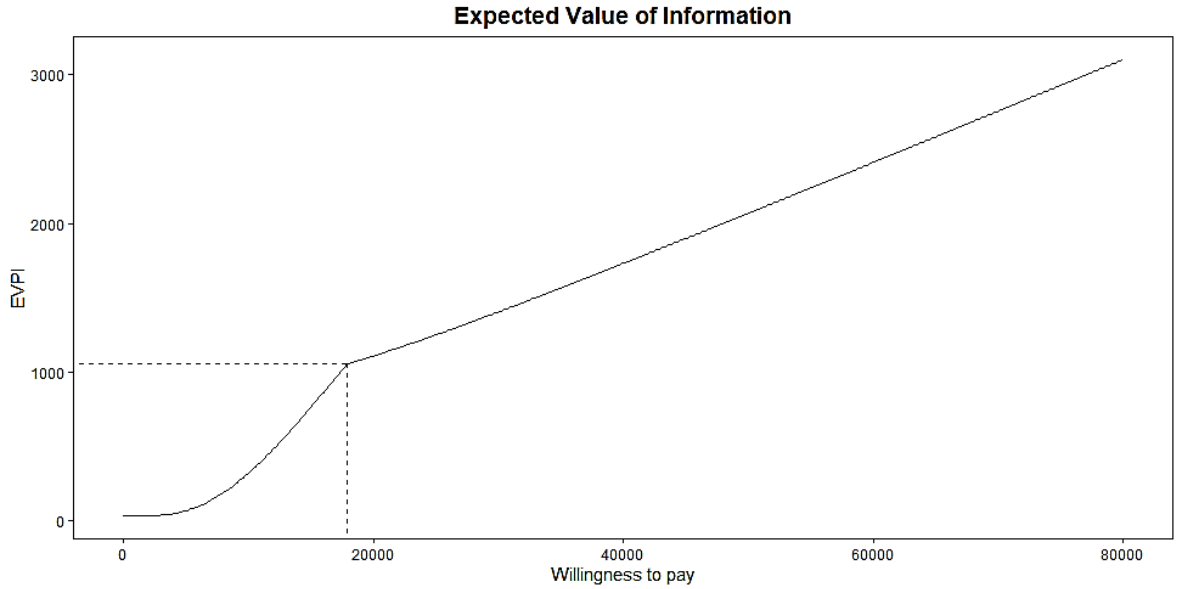

Figure 6 The EVPI for an individual patient in the base case PA

PA outcomes: scenarios for estimating the consequences of false-negative test results

It was reflected that all patients with a false-negative test result developed extrahepatic distant metastases before (delayed) treatment received **(=worst case)**, respectively, developed local/residual intrahepatic metastases only (i.e., no further non-local or distant tumour spread) before (delayed) treatment received **(=best case)**. Table 18 shows the results.

It can be seen that the consequences of the extent of tumour spread before (delayed) treatment received in the consecutive patient management were minimal. As a result, its impact on lifetime health and economic outcomes was negligible. The ICER estimates (€/QALY) fluctuated negligibly between the worst case and best case scenario (range from €23,188.22 to €23,902.68 saved per QALY lost) resulting in estimates which turned out to be strictly larger than the Dutch WTP threshold of €20,000/QALY for conditions with an intermediate disease burden.

Table 18 Worst and best case simulation results when varying the tumour spread before (delayed) treatment was received

	Intervention group		**Comparator group**	
	Number of events [1] (%)		*Number of events [1] (%)*	
Description	Worst case	Best case	Worst case	Best case
No recurrence observed in first-line follow-up (FU1)	3433 (34)	3438 (34)	3777 (38)	3786 (38)
Residual/local recurrence observed in FU1	1507 (15)	1516 (15)	885 (9)	885 (9)
Any intrahepatic non-local recurrence observed in FU1	1993 (20)	1996 (20)	1991 (20)	1997 (20)
Any extrahepatic distant recurrence observed in FU1	3067 (31)	3050 (31)	3347 (33)	3332 (33)
True test result in FU1	9183 (92)	9184 (92)	8581 (86)	8583 (86)
False-negative test result in FU1	752 (7)	750 (7)	1375 (14)	1373 (14)
False-positive test result in FU1	65 (1)	66 (1)	44 (0)	44 (0)
No downstream treatment received [2]	*457 (5)*	*459 (5)*	*407 (4)*	*410 (4)*
Downstream rLAT received	**4627 (46)**	**4622 (46)**	**4324 (43)**	**4321 (43)**
Downstream systemic therapy received	**4916 (49)**	**4919 (49)**	**5269 (53)**	**5269 (53)**
Below, the summarised lifetime health and economic consequences				
Costs	€39,019.20	€39,045.57	€40,241.94	€40,267.52
QALYs	**3.35**	**3.35**	**3.40**	**3.40**
LYGs	**5.11**	**5.11**	**4.99**	**5.00**
Resulting ICERs	Worst case		Best case	
ICER *(€/QALY)*	intervention group non-dominant **€23,902.68 saved per QALY lost**		intervention group non-dominant **€23,188.22 saved per QALY lost**	
ICER *(€/LYG)*	intervention group dominant **€10,481.23 saved per LYG**		intervention group dominant **€10,505.91 saved per LYG**	

[1] The mean number of events when simulating 10,000 runs of 10,000 individual patients with CRLM; [2] It should not be allowed that the intervention and comparator group influence the mean number of patients developing no recurrence resulting in an undesirable varying number of patients receiving no downstream treatment (elaborated in section 'Discussion').

PA outcomes: scenarios for estimating the consequences of prolonged survival if LAT received

The first scenario reflected a prolonged survival time of 12 months if rLAT was received (Table 19). It demonstrates that the health outcomes increased considerably (e.g., the QALYs gained increased by 0.28 and 0.34 for the intervention and comparator group respectively). Consequently, the ICER values also increased from €17,850.12 to €18,513.09 saved per QALY lost.

Another scenario reflected the consequences of applying gamma-distributed time-to-event data defined by the parameters *shape=3.20* and *rate=0.074*, instead of applying the log-logistic distribution, for estimating the overall survival times if rLAT was received (Table 20). It demonstrates that the

distribution applied substantially affected the ICER. The ICER estimates (€/QALY) fluctuated considerably between the scenarios (range from €17,850.12 to €27,718.13 saved per QALY lost) resulting in estimates which turned out to be either larger or smaller than the Dutch WTP threshold of €20,000/QALY for conditions with an intermediate disease burden.

Table 19 Costs, QALYs and LYGs per strategy resulting in the new ICER per QALY and LYG respectively when a prolonged survival time of 12 months was rewarded if rLAT was received

Strategy	**Costs**	**QALYs**	**LYGs**	**ICER** *(€/QALY)*	**ICER** *(€/LYG)*
Intervention group [1]	€38,924.47 (€39,254.01)	3.66 (3.38)	5.56 (5.08)	intervention group non-dominant **€18,513.09 saved per QALY lost** (€17,850.12)	intervention group dominant **€15,628.60 saved per LYG** (€14,668.67)
Comparator group [1]	€40,798.91 (€40,681.57)	3.76 (3.42)	5.44 (4.98)		

[1] For smooth comparison, the original health outcomes of the base case are shown in brackets

Table 20 Costs, QALYs and LYGs per strategy resulting in the new ICER per QALY and LYG respectively when gamma-distributed time-to-event data was assigned to the estimated overall survival times if rLAT was received

Strategy	**Costs**	**QALYs**	**LYGs**	**ICER** *(€/QALY)*	**ICER** *(€/LYG)*
Intervention group [1]	€39,272.40 (€39,254.01)	3.32 (3.38)	5.02 (5.08)	intervention group non-dominant **€27,718.13 saved per QALY lost** (€17,850.12)	intervention group dominant **€12,039.72 saved per LYG** (€14,668.67)
Comparator group [1]	€40,648.37 (€40,681.57)	3.37 (3.42)	4.91 (4.98)		

[1] For smooth comparison, the original health outcomes of the base case are shown in brackets

In addition, applying gamma-distributed time-to-event data for estimating the overall survival times if rLAT was received, affected the uncertainty surrounding the cost-effectiveness estimates (Figure 7). Clearly, the number of extreme values decreased so that the unique point estimates with extreme values such as *2 QALYs lost*, respectively, *3 QALYs lost* (while saving of costs) no longer occurred.

Figure 7 Applying gamma-distributed time-to-event data for estimating the overall survival times if rLAT was received. Estimates below the diagonal represent simulations in which the intervention group was cost-effective (WTP €20,000/QALY)

Conclusions

This economic evaluation estimated the cost-effectiveness of adopting ^{18}F-FDG PET/CT in first-line follow-up compared to conventional CECT workup. Adopting ^{18}F-FDG PET/CT resulted in a nearly doubled mean number of residual/local recurrences observed. Furthermore, the mean number of false-negative test results was nearly halved. Consequently, we observed a small increase in the number of LAT procedures provided when adopting ^{18}F-FDG PET/CT in first-line follow-up, such that we could waive (or postpone) providing less favourable complex systemic therapy in a meaningful number of patients.

However, adopting ^{18}F-FDG PET/CT was not proved to be a cost-effective alternative to conventional follow-up by CECT. Specifically, the uncertainty surrounding the cost-effectiveness estimates was demonstrated in the incremental cost-effectiveness planes (Figures 4,7). In addition, conclusions about our findings regarding lifetime health and economic consequences should be made with caution as the results of the scenario analyses did not prove to be quite robust. Accordingly, ICER estimates (€/QALY) fluctuated considerably between the discussed scenarios.

Discussion

Complexity in the CRLM care pathway regarding LAT, systemic therapy and multiple lines follow-up diagnostics is unlikely to be resolved by randomized controlled clinical trials only, indicating an opportunity for DES modelling utilizing observational data to provide information on optimal intervention and diagnostics sequencing (123–125). The ever-increasing pressure to ensure the most efficient and effective use of limited health service resources will also encourage clinicians and policy makers to make use of modelling solutions (92). From a policy perspective, the value of model-based analysis lies not simply in its ability to generate a precise point estimate for a specified health outcome, but also in the systematic examination and responsible reporting of uncertainty surrounding the ultimate clinical decision-making being addressed (89,126).

Despite the estimated mean number of false-negative test results that was nearly halved, adopting ^{18}F-FDG PET/CT in routine first-line follow-up was not proven to be cost-effective over conventional follow-up by CECT. Hence, our health-economic simulation results are in line with the overall conclusions of previous cost-effectiveness studies published in which compelling cost-effectiveness outcomes are also absent (8,10). There are several reasons for this absence but it all comes down to the quality and quantity of observational data available resulting in a large uncertainty surrounding the parameter values used for simulation. Most importantly, the aforementioned studies had to deal with a small study population (n<50) while no standard biopsy was taken of suspected tumour lesions **(1)**. Also, there is much uncertainty surrounding survival rates, quality of life and the assigned costs regarding

sequentially received downstream treatment procedures **(2)**. Moreover, there is little evidence available on CRLM progression rates over time **(3)**. And, the consequences of individual lesion-based false test results without a disease state shift at patient-level remain unclear **(4)**.

Moreover, the existing literature on the cost-effectiveness of adopting ^{18}F-FDG PET/CT in routine first-line follow-up is very limited and has taken different approaches in terms of modelling methods, model structures and the time horizon of simulation. Consequently, this results in considerably different cost-effectiveness point estimates. Because of the lack of model validation and the wide variation in structural assumptions, the exact cost-effectiveness of adopting ^{18}F-FDG PET/CT in routine first-line follow-up remains highly uncertain in current literature.

Clear is that the improved accuracy of first-line follow-up when adopting ^{18}F-FDG PET/CT strongly underlines the incremental interest and benefits of minimally invasive TA therapy. The increased sensitivity resulted in the detection of smaller tumours which are more amendable to rLAT, and, which can postpone complex systemic therapy. Accordingly, in the existing literature on the clinical use of ^{18}F-FDG PET/CT, it is suggested that the use of ^{18}F-FDG PET/CT can lead to a change in the primarily intended patient management in up to 20% of all CRLM patients. (5,10,16,70,73,84–101)

Our cost-effectiveness study has several limitations that may restrict the interpretation of our simulation results for daily clinical practice. First of all, our DES model is a simplified reflection of clinical practice and the accuracy of the estimated probabilities, particular time-to-event data, costs and utilities are dependent on the availability and quality of representative literature source data. Note that this is a limitation of all model-based cost-effectiveness analyses.

The overall survival curves of the reported downstream treatment strategies and the progression-free survival curves of higher-order follow-up were based on such literature evidence (12,19,29,60). The survival curves were digitized for particular time points and the number at risk at the beginning of each interval was calculated. The reliability of the reconstructed and digitized data depended on two related elements: the quality of the initial input **(1)** and the level of information provided by the publication **(2)**. (103) Noteworthy, we experienced implementation struggles while undertaking the initial digitization and pre-processing of the survival data via multiple open-source digitizing software.

Furthermore, the inability to derive separate survival curves for different subgroups or to model the joint effects of covariates and treatment with the algorithm of *Guyot et al.* might affect the estimated treatment effects due to aggregation bias. Aggregation over the covariate tends to bias the treatment effects towards the null, and the extent of the bias increases with the strength of the covariate effect. (103,144) In general, biases can significantly affect health economic outcome estimates and should be minimised. Typically, there is a substantial difference between theoretical study results and real-life observations due to bias and confounding factors. (145)

Therefore, we extensively compared the reconstructed survival data with the results reported in the original publication (e.g., survival probabilities, medians, hazard ratios). The reported outcomes were somewhat striking since the digitized survival times of systemic therapy were not strongly differentiating from the survival times of rLAT. Despite the reported inclusion and exclusion criteria, it proved difficult to determine whether the study populations of the studies consulted were entirely appropriate comparators. It also remained unclear whether unconscious selection bias occurred within the consulted studies.

Second, the magnitude of uncertainty when extrapolating survival curves may be substantial. This is an issue for all time-to-event studies where extrapolation is unavoidable. In literature, differences in estimates of uncertainty are collectively observed even when models provided near-identical point estimates. (97–102)

Third, applied was a model structure that was process-oriented and treatment-driven incorporating a detailed colorectal liver disease-specific clinical care pathway. Adopting a treatment-driven structure was preferable over a health-state-driven structure to reflect the complex pathway and the dynamics of clinical practice more realistically (123,124). However, reflecting multiple treatment lines with varying imaging modalities, different adverse events and their impact on health outcomes while considering downstream effects of patients' treatment history, proved to require a substantial number of model states and assumptions regarding the occurrence of CRLM recurrence, that we decided to average the cost and health outcomes of downstream treatments. As a consequence of our model focusing on a single LAT procedure with multiple lines of diagnostic follow-up while only averaging the consequences of downstream treatment strategies, there was a profound source of structural uncertainty implying the cost-effectiveness of multiple sequential treatment strategies across the CRLM care pathway remained highly uncertain.

Fourth, when classifying all unique and raw observations on both imaging modalities from the multidisciplinary evaluation of first-line follow-up diagnostics to a set of four mutually exclusive and collectively exhaustive disease states, loss of information was unavoidable (i.e., a unique patient with intrahepatic and extrahepatic tumour lesions was classified to the same disease state as another patient with extrahepatic tumour lesions only).

In particular, we must be very careful when interpreting false-negative test results. For all disease states where CRLM recurrence was correctly indicated, we may overlook individual tumour lesions whereas it would not be noticed as a false-negative test result by our model. The reason is that the incorporated evidence was not sufficient to distinguish lesion-based analysis in our patient-level model (e.g., you need to know the probability of a negative test result for an individual tumour cell within the group of patients with proven intrahepatic respectively extrahepatic CRLM, as well as the probability of a positive test result for an individual tumour cell within these disease states).

Consequently, we argue that the true added value of adopting ^{18}F-FDG PET/CT in first-line follow-up is systematically underestimated. On the other hand, we have to consider that false-negative test results are about missing the smaller and less dangerous tumour cells. Moreover, some tumour cells (i.e., micro cancer cells) will never be detected by any imaging modality. But for convenience, our model assumed that all tumour cells are detected after finishing the total follow-up duration (a period of at most twelve years). So that after multiple lines of higher-order follow-up no single tumour cell remained undetected.

Further, note that after the multidisciplinary evaluation of the first-line follow-up diagnostics, a second view of the same two observers appointed from the expert panel was realized and consensus was reached if any disagreement occurred between two assessments of the same imaging modality. In the case of assessing the ^{18}F-FDG PET/CT scans, no inter-observer variability was shown. But the assessments of the conventional CECT scans showed inter-observer variability and consensus was reached for a dozen CECT scans. Typically, in all cases reaching consensus meant a repair of a false-negative to a true positive test result. In daily clinical practice, however, a single expert is assessing the CECT scans and such a repair would not be possible. Therefore, we argue that our simulation study is underestimating the number of false-negative CECT scans compared to what would be expected in daily clinical practice.

For all disease states where patient-level recurrence was correctly indicated, false-positive individual tumour lesions could also arise. Again, the reason is that incorporated evidence was not sufficient to distinguish lesion-based analysis in our patient-level model. Dependent on the type of downstream patient management, false-positive individual tumour lesions within the correctly indicated disease state might introduce unnecessary operating costs to the healthcare system and are burdensome to the patient. In the case of systemic therapy, probably no extra costs or health burden on patients was introduced (i.e., other true positive tumour lesions still benefit from the therapy) but if LAT was provided unnecessary operating costs and health burden on patients were introduced (i.e., LAT may be received for a segment of purely healthy liver tissue).

Fifth, in literature, reported HR-QoLs vary markedly between studies and this adds to the uncertainty regarding our cost-effectiveness estimates. Despite several publications about CRLM-related HR-QoLs, there did not exist a recent publication that was methodologically robust with a full range of values for all health states of our interest. (111) Therefore, health outcomes were also expressed in LYGs.

Sixth, due to the absence of biopsies taken of suspected tumour lesions in the hospital data, we used aggregated literature evidence on the diagnostic performance of the imaging modalities consulted. The literature-based estimated performance projected on the hospital data resulted in a biased a priori probability to develop no recurrence over a lifetime horizon when a patient is assigned to the intervention group that is, in the base case, 0.4% higher than in the comparator group. Consequently, there existed a biased a priori probability to develop CRLM recurrence over a lifetime horizon when a patient is assigned to the intervention group that is, in the base case, 0.4% lower than in the comparator group.

This small difference originated because we combined aggregated literature evidence with patient-level hospital data available and can be solved by minimally correcting for the ratios of false-negative and false-positive test results.

Seventh, we overestimated the mean value of the sensitivity of conventional first-line follow-up by CECT used for simulation (Table 10) compared to the mean sensitivity value of the asymmetrically distributed evidence from literature (Table 9) by almost 4%. Even more striking was that the estimated median values of the sensitivity and specificity of both imaging strategies were unrealistically high and, therefore, almost completely non-distinguishing. As a consequence of the overestimated mean value of the sensitivity of conventional CECT **(1)** and the extremely skewed distribution of the sensitivity and specificity of both imaging strategies **(2)** used for simulation, we overestimated the diagnostic performance of the comparator group relatively to the intervention group and compared to what would be realistic for clinical practice. Finally, the chance that the intervention group would be accepted as standard care decreased drastically.

Further, it should be emphasized that the discussed EVPI represents a maximum value of additional research. As long as the cost of a given research project is less than the EVPI, there is at least a potential of representing efficient use of resources. To decide if further research will be worthwhile and identify an efficient research design, we need to consider the marginal benefits and marginal cost of sample information. VOI analysis can be extended to establish the expected value of sample information for particular research designs and to compare these expected benefits of research to the expected costs. This type of analysis provides a societal payoff to alternative designs and can be used to establish optimal sample size, optimal allocation of patients within a clinical trial and appropriate follow-up workup. (122)

Lastly, it should be recognized that some interesting subjects were out of the scope of our study. We did not consider the patient's preference and adherence to treatment procedures and diagnostic follow-up. Also, we did not consider the hospital's capacity allocation and potential waiting lists. Furthermore, during the clinical trial, we did not report on any comorbidities or secondary cancer diseases presented.

We wrap up by addressing the use of observational data as applied in our simulation study. In literature, observational studies based on real-world data have been recognized as a valuable source of evidence for comparative cost-effectiveness research in oncology. In particular, when evidence from randomized controlled trials is not available **(1)** or when results from these trials only apply to a selected group of patients **(2)**. Although observational studies are subject to selection bias, careful analysis of real-world data has the potential to provide valuable information on the comparative cost-effectiveness of different diagnostics. However, regardless of whether clinical outcomes of different patient management are studied through clinical trials or observational studies, it remains highly challenging to translate results from these studies to an optimal care pathway that informs patient management across multiple downstream lines of interventions. (123–125) Moreover, it is argued that the hospital's diagnostic

surveillance schedules are most intensive during the first three years than at later time points, but there is no consensus on the optimal surveillance interval. (74,127,130,146,147) This strongly contributes to the uncertainty surrounding the ultimate downstream patient management being addressed.

Implications for clinical practice and further research

Model-based analysis is an unavoidable fact of life in health economic evaluations (148,149). Our DES modelling study contributes to a scientific foundation on how to reach a cost-effective follow-up care pathway for patients with CRLM. This study provided insights into the benefits of adopting ^{18}F-FDG PET/CT in routine first-line follow-up under various scenarios without intervening in the real process. Keeping this in mind, this study served as a bridge between clinical practice and mathematical modelling. It supported clinical decision-making by simulation modelling and analysis in which we used computer technology to estimate time-to-events, which is similar to what we would do in a randomized controlled clinical trial.

This explorative study could be seen as the first step for further clinical research regarding the added value of ^{18}F-FDG PET/CT in the restaging of CRLM. This study could potentially influence daily clinical practice, but further investigations should be performed to determine the cost-effectiveness when comprehensively incorporating multiple lines of sequentially provided downstream treatment strategies.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.
2. Kuipers EJ, Grady W, Lieberman D. Colorectal Cancer. Nat Rev Dis Prim. 2015;
3. Brenner H, Kloor M, Pox CP. Colorectal cancer. Lancet. 2014;383(9927):1490–502.
4. Van Cutsem E, Cervantes A, Adam R, Sobrero A, Van Krieken JH, Aderka D, et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. Ann Oncol. 2016;27(8):1386–422.
5. Choi SH, Kim SY, Park SH, Kim KW, Lee JY, Lee SS, et al. Diagnostic performance of CT, gadoxetate disodium-enhanced MRI, and PET/CT for the diagnosis of colorectal liver metastasis: Systematic review and meta-analysis. J Magn Reson Imaging. 2018;47(5):1237–50.
6. Roberts KJ, Sutton AJ, Prasad KR, Toogood GJ, Lodge JPA. Cost-utility analysis of operative versus non-operative treatment for colorectal liver metastases. Br J Surg. 2015;102(4):388–98.
7. Froelich MF, Schnitzer ML, Rathmann N, Tollens F, Unterrainer M, Rennebaum S, et al. Cost-effectiveness analysis of local ablation and surgery for liver metastases of oligometastatic colorectal cancer. Cancers (Basel). 2021;13(7):1–11.
8. Chen W, Zhuang H, Cheng G, Torigian DA, Alavi A. Comparison of FDG-PET, MRI and CT for post radiofrequency ablation evaluation of hepatic tumours. Ann Nucl Med. 2013;27(1):58–64.

9. Puijk RS, Ruarus AH, Vroomen LGPH, van Tilborg AAJM, Scheffer HJ, Nielsen K, et al. Colorectal liver metastases: Surgery versus thermal ablation (COLLISION) - a phase III single-blind prospective randomized controlled trial. BMC Cancer. 2018;18(1):1–13.

10. Schnitzer ML, Froelich MF, Gassert FG, Huber T, Gresser E, Schwarze V, et al. Contrast-Enhanced CT after Ablation of Liver Metastases of Colorectal Carcinoma — A Cost-E ff ectiveness Analysis.

11. Engstrand J, Nilsson H, Strömberg C, Jonas E, Freedman J. Colorectal cancer liver metastases - a population-based study on incidence, management and survival. BMC Cancer. 2018;18(1):1–11.

12. Ruers T, Van Coevorden F, Punt CJA, Pierie JPEN, Borel-Rinkes I, Ledermann JA, et al. Local Treatment of Unresectable Colorectal Liver Metastases: Results of a Randomized Phase II Trial. J Natl Cancer Inst. 2017;109(9):1–10.

13. van Amerongen MJ, Jenniskens SFM, van den Boezem PB, Fütterer JJ, de Wilt JHW. Radiofrequency ablation compared to surgical resection for curative treatment of patients with colorectal liver metastases – a meta-analysis. Hpb [Internet]. 2017;19(9):749–56. Available from: http://dx.doi.org/10.1016/j.hpb.2017.05.011

14. De Ridder JAM, Van Der Stok EP, Mekenkamp LJ, Wiering B, Koopman M, Punt CJA, et al. Management of liver metastases in colorectal cancer patients: A retrospective case-control study of systemic therapy versus liver resection. Eur J Cancer [Internet]. 2016;59:13–21. Available from: http://dx.doi.org/10.1016/j.ejca.2016.02.003

15. Samim M, Molenaar IQ, Seesing MFJ, van Rossum PSN, van den Bosch MAAJ, Ruers TJM, et al. The diagnostic performance of 18F-FDG PET/CT, CT and MRI in the treatment evaluation of ablation therapy for colorectal liver metastases: A systematic review and meta-analysis. Surg Oncol [Internet]. 2017;26(1):37–45. Available from: http://dx.doi.org/10.1016/j.suronc.2016.12.006

16. Yang G, Wang G, Sun J, Xiong Y, Li W, Tang T, et al. The prognosis of radiofrequency ablation versus hepatic resection for patients with colorectal liver metastases: A systematic review and meta-analysis based on 22 studies. Int J Surg [Internet]. 2021;87(November 2020):105896. Available from: https://doi.org/10.1016/j.ijsu.2021.105896

17. Lin YM, Paolucci I, Brock KK, Odisio BC. Image-guided ablation for colorectal liver metastasis: Principles, current evidence, and the path forward. Cancers (Basel). 2021;13(16):1–21.

18. van der Geest LGM, Lam-Boer J, Koopman M, Verhoef C, Elferink MAG, de Wilt JHW. Nationwide trends in incidence, treatment and survival of colorectal cancer patients with synchronous metastases. Clin Exp Metastasis. 2015;32(5):457–65.

19. Ruers T, Punt C, Van coevorden F, Pierie JPEN, Borel-Rinkes I, Ledermann JA, et al. Radiofrequency ablation combined with systemic treatment versus systemic treatment alone in patients with non-resectable colorectal liver metastases: A randomized eortc intergroup phase ii study (EORTC 40004). Ann Oncol [Internet]. 2012;23(10):2619–26. Available from: https://doi.org/10.1093/annonc/mds053

20. Morris EJA, Forman D, Thomas JD, Quirke P, Taylor EF, Fairley L, et al. Surgical management and outcomes of colorectal cancer liver metastases. Br J Surg. 2010;97(7):1110–8.

21. Serrano PE, Gu CS, Husien M, Jalink D, Ritter A, Martel G, et al. Risk factors for survival following recurrence after first liver resection for colorectal cancer liver metastases. J Surg Oncol. 2019;120(8):1420–6.

22. Spelt L, Andersson B, Nilsson J, Andersson R. Prognostic models for outcome following liver resection for colorectal cancer metastases: A systematic review. Eur J Surg Oncol [Internet]. 2012;38(1):16–24. Available from: http://dx.doi.org/10.1016/j.ejso.2011.10.013

23. Guckenberger M, Lievens Y, Bouma AB, Collette L, Dekker A, deSouza NM, et al. Characterisation and classification of oligometastatic disease: a European Society for Radiotherapy and Oncology and European Organisation for Research and Treatment of Cancer consensus recommendation. Lancet Oncol [Internet]. 2020;21(1):e18–28. Available from: http://dx.doi.org/10.1016/S1470-2045(19)30718-1

24. Rumpold H, Niedersüß-Beke D, Heiler C, Falch D, Wundsam HV, Metz-Gercek S, et al. Prediction of mortality in metastatic colorectal cancer in a real-life population: a multicenter explorative analysis. BMC Cancer. 2020;20(1):1–9.

25. Boysen AK, Spindler KL, Høyer M, Mortensen FV, Christensen TD, Farkas DK, et al. Metastasis directed therapy for liver and lung metastases from colorectal cancer—A population-based study. Int J Cancer. 2018;143(12):3218–26.

26. Leung U, Gönen M, Allen PJ. Colorectal Cancer Liver Metastases and Concurrent Extrahepatic Disease Treated with Resection. Ann Surg. 2017;265(1):158–65.

27. Kishiki T, Lapin B, Matsuoka H, Watanabe T, Takayasu K, Kojima K, et al. Optimal Surveillance Protocols after Curative Resection in Patients with Stage IV Colorectal Cancer: A Multicenter Retrospective Study. Dis Colon Rectum. 2018;61(1):51–7.

28. Dai S, Ye Y, Kong X, Li J, Ding K. A predictive model for early recurrence of colorectal-cancer liver metastases based on clinical parameters. Gastroenterol Rep. 2021;9(3):241–51.

29. Dijkstra M, Nieuwenhuizen S, Puijk RS, Timmer FEF, Geboers B, Schouten EAC, et al. Thermal ablation compared to partial hepatectomy for recurrent colorectal liver metastases: An amsterdam colorectal liver met registry (amcore) based study. Cancers (Basel). 2021;13(11).

30. Elfrink AKE, Nieuwenhuizen S, van den Tol MP, Burgmans MC, Prevoo W, Coolsen MME, et al. Hospital variation in combined liver resection and thermal ablation for colorectal liver metastases and impact on short-term postoperative outcomes: a nationwide population-based study. Hpb. 2021;23(6):827–39.

31. Nieuwenhuizen S, Puijk RS, van den Bemd B, Aldrighetti L, Arntz M, van den Boezem PB, et al. Resectability and ablatability criteria for the treatment of liver only colorectal metastases: Multidisciplinary consensus document from the COLLISION trial group. Cancers (Basel). 2020;12(7):1–17.

32. Scherman P, Syk I, Holmberg E, Naredi P, Rizell M. Influence of primary tumour and patient factors on survival in patients undergoing curative resection and treatment for liver metastases from colorectal cancer. BJS open. 2020;4(1):118–32.

33. Scherman P, Syk I, Holmberg E, Naredi P, Rizell M. Impact of patient, primary tumour and metastatic pattern including tumour location on survival in patients undergoing ablation or resection for colorectal liver metastases: A population-based national cohort study. Eur J Surg Oncol [Internet]. 2021;47(2):375–83. Available from: https://doi.org/10.1016/j.ejso.2020.07.030

34. Liu W, Liu JM, Wang K, Wang HW, Xing BC. Recurrent colorectal liver metastasis patients could benefit from repeat hepatic resection. BMC Surg [Internet]. 2021;21(1):1–10. Available from: https://doi.org/10.1186/s12893-021-01323-y

35. Gillams A, Goldberg N, Ahmed M, Bale R, Breen D, Callstrom M, et al. Thermal ablation of colorectal liver metastases: a position paper by an international panel of ablation experts, the interventional oncology sans frontières meeting 2013. Eur Radiol. 2015;25(12):3438–54.

36. De Baere T, Tselikas L, Pearson E, Yevitch S, Boige V, Malka D, et al. Interventional oncology for liver and lung metastases from colorectal cancer: The current state of the art. Diagn Interv Imaging. 2015;96(6):647–54.

37. Rosati G, Ambrosini G, Barni S, Andreoni B, Corradini G, Luchena G, et al. A randomized trial of intensive versus minimal surveillance of patients with resected Dukes B2-C colorectal carcinoma. Ann Oncol [Internet]. 2016;27(2):274–80. Available from: https://doi.org/10.1093/annonc/mdv541

38. Lin J, Peng J, Zhao Y, Luo B, Zhao Y, Deng Y, et al. Early recurrence in patients undergoing curative resection of colorectal liver oligometastases: identification of its clinical characteristics, risk factors, and prognosis. J Cancer Res Clin Oncol. 2018;144(2):359–69.

39. Vauthey JN, Kawaguchi Y. Innovation and Future Perspectives in the Treatment of Colorectal Liver Metastases. J Gastrointest Surg. 2020;24(2):492–6.

40. Kanas G, Taylor A, Kelsh MA, Mowat FS, Alexander DD, Choti MA, et al. Survival after liver resection in metastatic colorectal cancer : review and meta-analysis of prognostic factors. Clin Epidemiol. 2012;283–301.

41. Fong Y, Fortner J, Sun RL, Brennan MF, Blumgart LH. Clinical Score for Predicting Recurrence After Hepatic Resection for Metastatic Colorectal Cancer. Ann Surg [Internet]. 1999;230(3):309. Available

from:http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1420876/%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC1420876/pdf/19990900s00004p309.pdf

42. Sasaki K, Morioka D, Conci S, Margonis GA, Sawada Y, Ruzzenente A, et al. The Tumour Burden Score: A New "metro-ticket" Prognostic Tool for Colorectal Liver Metastases Based on Tumour Size and Number of Tumours. Ann Surg. 2018;267(1):132–41.

43. Tinguely P, Dal G, Bottai M, Nilsson H, Freedman J, Engstrand J. Microwave ablation versus resection for colorectal cancer liver metastases – A propensity score analysis from a population-based nationwide registry. Eur J Surg Oncol [Internet]. 2020;46(3):476–85. Available from: https://doi.org/10.1016/j.ejso.2019.12.002

44. Chen Y, Chang W, Ren L, Chen J, Tang W, Liu T, et al. Comprehensive Evaluation of Relapse Risk (CERR) Score for Colorectal Liver Metastases: Development and Validation. Oncologist. 2020;25(7):e1031–41.

45. Frühling P, Urdzik J, Strömberg C, Isaksson B. Composite Score: prognostic tool to predict survival in patients undergoing surgery for colorectal liver metastases. BJS open. 2021;5(5).

46. Sasaki K, Margonis GA, Andreatos N, Zhang XF, Buettner S, Wang J, et al. The prognostic utility of the "Tumour Burden Score" based on preoperative radiographic features of colorectal liver metastases. J Surg Oncol. 2017;116(4):515–23.

47. Wimmer K, Schwarz C, Szabo C, Bodingbauer M, Tamandl D, Mittlböck M, et al. Impact of Neoadjuvant Chemotherapy on Clinical Risk Scores and Survival in Patients with Colorectal Liver Metastases. Ann Surg Oncol. 2017;24(1):236–43.

48. Wu Y, Guo T, Xu Z, Liu F, Cai S, Wang L, et al. Risk scoring system for recurrence after simultaneous resection of colorectal cancer liver metastasis. Ann Transl Med. 2021;9(12):966–966.

49. Guinney J, Dienstmann R, Wang X, De Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350–6.

50. Margonis GA, Sasaki K, Gholami S, Kim Y, Andreatos N, Rezaee N. The Genetic And Morphological Evoluation (GAME) score for patients with Colorectal Liver Metastases. Br J Surg. 2018;105(9):1210–20.

51. Zhao B, Gabriel RA, Vaida F. Predicting Overall Survival in Patients with Metastatic Rectal Cancer: a Machine Learning Approach. J Gastrointest Surg. 2020;24:1165–72.

52. Viganò L, Pedicini V, Comito T, Carnaghi C, Costa G, Poretti D, et al. Aggressive and Multidisciplinary Local Approach to Iterative Recurrences of Colorectal Liver Metastases. World J Surg. 2018;42(8):2651–9.

53. Hellingman T, Kuiper BI, Buffart LM, Meijerink MR, Versteeg KS, Swijnenburg RJ, et al. Survival Benefit of Repeat Local Treatment in Patients Suffering From Early Recurrence of Colorectal Cancer Liver Metastases. Clin Colorectal Cancer [Internet]. 2021;20(4):e263–72. Available from: https://doi.org/10.1016/j.clcc.2021.07.007

54. Tai K, Komatsu S, Sofue K, Kido M, Tanaka M, Kuramitsu K, et al. Total tumour volume as a prognostic factor in patients with resectable colorectal cancer liver metastases. BJS open. 2020;4(3):456–66.

55. de Baere T, Tselikas L, Yevich S, Boige V, Deschamps F, Ducreux M, et al. The role of image-guided therapy in the management of colorectal cancer metastatic disease. Eur J Cancer. 2017;75:231–42.

56. Wong SL, Mangu PB, Choti MA, Crocenzi TS, Dodd GD, Dorfman GS, et al. American Society of Clinical Oncology 2009 clinical evidence review on radiofrequency ablation of hepatic metastases from colorectal cancer. J Clin Oncol. 2010;28(3):493–508.

57. Hompes D, Prevoo W, Ruers T. Radiofrequency ablation as a treatment tool for liver metastases of colorectal origin. Cancer Imaging. 2011;11(1):23–30.

58. Chiappa A, Bertani E, Zbar AP, Foschi D, Fazio N, Zampino M, et al. Optimizing treatment of hepatic metastases from colorectal cancer: Resection or resection plus ablation? Int J Oncol. 2016;48(3):1280–9.

59. Nielsen K, Scheffer HJ, Volders JH, van der Vorst MJDL, van Tilborg AAJM, Comans EFI, et al. Radiofrequency Ablation to Improve Survival After Conversion Chemotherapy for Colorectal Liver Metastases. World J Surg. 2016;40(8):1951–8.

60. Nieuwenhuizen S, Dijkstra M, Puijk RS, Timmer FEF, Nota IM, Opperman J, et al. Thermal ablation versus stereotactic ablative body radiotherapy to treat unresectable colorectal liver metastases: A comparative analysis from the prospective amsterdam core registry. Cancers (Basel). 2021;13(17).

61. Dupré A, Jones RP, Diaz-Nieto R, Fenwick SW, Poston GJ, Malik HZ. Curative-intent treatment of recurrent colorectal liver metastases: A comparison between ablation and resection. Eur J Surg Oncol [Internet]. 2017 Oct 1 [cited 2022 May 12];43(10):1901–7. Available from: http://dx.doi.org/10.1016/j.ejso.2017.08.008

62. Wang L-J, Zhang Z-Y, Yan X-L, Yang W, Yan K, Xing B-C. Radiofrequency ablation versus resection for technically resectable colorectal liver metastasis: a propensity score analysis. [cited 2022 May 12]; Available from: https://doi.org/10.1186/s12957-018-1494-3

63. Winkelmann MT, Clasen S, Pereira PL, Hoffmann R. Local treatment of oligometastatic disease: Current role. Br J Radiol. 2019;92(1100).

64. Decadt B, Siriwardena AK. Radiofrequency ablation of liver tumours: Systematic review. Lancet Oncol. 2004;5(9):550–60.

65. Kron P, Linecker M, Jones RP, Toogood GJ, Clavien PA, Lodge JPA. Ablation or Resection for Colorectal Liver Metastases? A Systematic Review of the Literature. Front Oncol. 2019;9(October).

66. Di Martino M, Rompianesi G, Mora-Guzmán I, Martín-Pérez E, Montalti R, Troisi RI. Systematic review and meta-analysis of local ablative therapies for resectable colorectal liver metastases. Eur J Surg Oncol. 2020;46(5):772–81.

67. Takahashi H, Kahramangil B, Kose E, Berber E. A comparison of microwave thermosphere versus radiofrequency thermal ablation in the treatment of colorectal liver metastases. Hpb [Internet]. 2018;20(12):1157–62. Available from: https://doi.org/10.1016/j.hpb.2018.05.012

68. De'Angelis N, Baldini C, Brustia R, Pessaux P, Sommacale D, Laurent A, et al. Surgical and regional treatments for colorectal cancer metastases in older patients: A systematic review and meta-analysis. PLoS One. 2020;15(4).

69. Filippiadis DK, Velonakis G, Kelekis A, Sofocleous CT. The role of percutaneous ablation in the management of colorectal cancer liver metastatic disease. Diagnostics. 2021;11(2):1–13.

70. Tanis E, Nordlinger B, Mauer M, Sorbye H, Van Coevorden F, Gruenberger T, et al. Local recurrence rates after radiofrequency ablation or resection of colorectal liver metastases. Analysis of the European Organisation for Research and Treatment of Cancer #40004 and #40983. Eur J Cancer [Internet]. 2014;50(5):912–9. Available from: http://dx.doi.org/10.1016/j.ejca.2013.12.008

71. Van Cutsem E, Verheul HMW, Flamen P, Rougier P, Beets-Tan R, Glynne-Jones R, et al. Imaging in colorectal cancer: Progress and challenges for the clinicians. Cancers (Basel). 2016;8(9).

72. Aarntzen EHJG, Heijmen L, Oyen WJG. 18F-FDG PET/CT in local ablative therapies: A systematic review. J Nucl Med. 2018;59(4):551–6.

73. Howard BA, Wong TZ. 18F-FDG-PET/CT Imaging for Gastrointestinal Malignancies. Radiol Clin North Am [Internet]. 2021;59(5):737–53. Available from: https://doi.org/10.1016/j.rcl.2021.06.001

74. Gerke O, Ehlers K, Motschall E, Høilund-Carlsen PF, Vach W. PET/CT-Based Response Evaluation in Cancer—a Systematic Review of Design Issues. Mol Imaging Biol. 2020;22(1):33–46.

75. Gauthé M, Richard-Molard M, Cacheux W, Michel P, Jouve JL, Mitry E, et al. Role of fluorine 18 fluorodeoxyglucose positron emission tomography/computed tomography in gastrointestinal cancers. Dig Liver Dis [Internet]. 2015;47(6):443–54. Available from: http://dx.doi.org/10.1016/j.dld.2015.02.005

76. Jayaprakasam VS, Paroder V, Schöder H. Variants and Pitfalls in PET/CT Imaging of Gastrointestinal Cancers. Semin Nucl Med. 2021;51(5):485–501.

77. Shim JR, Lee SD, Han SS, Lee SJ, Lee DE, Kim SK, et al. Prognostic significance of 18F-FDG PET/CT

in patients with colorectal cancer liver metastases after hepatectomy. Eur J Surg Oncol. 2018;44(5):670–6.

78. Samim M, Prevoo W, de Wit-van der Veen BJ, Kuhlmann KF, Ruers T, van Hillegersberg R, et al. 18F-FDG PET as novel imaging biomarker for disease progression after ablation therapy in colorectal liver metastases. Eur J Nucl Med Mol Imaging. 2017;44(7):1165–75.

79. Nakajo M, Kajiya Y, Tani A, Jinguji M, Nakajo M, Kitazono M, et al. A pilot study for texture analysis of 18F-FDG and 18F-FLT-PET/CT to predict tumour recurrence of patients with colorectal cancer who received surgery. Eur J Nucl Med Mol Imaging. 2017;44(13):2158–68.

80. Samim M, El-Haddad GE, Molenaar IQ, Prevoo W, Van Den Bosch MAAJ, Alavi A, et al. [18F]fluorodeoxyglucose PET for interventional oncology in liver malignancy. PET Clin. 2014;9(4):469–95.

81. Bonichon F, Godbert Y, Gangi A, Buy X, Palussière J. PET/computed tomography and thermoablation (radiofrequency, microwave, cryotherapy, laser interstitial thermal therapy). PET Clin [Internet]. 2015;10(4):519–40. Available from: http://dx.doi.org/10.1016/j.cpet.2015.05.008

82. Rice SR, Chuong M, Koroulakis A, Siddiqui OM, Sharma AM, Simone CB, et al. The Utility of PET/Computed Tomography for Radiation Oncology Planning, Surveillance, and Prognosis Prediction of Gastrointestinal Tumours. PET Clin [Internet]. 2020;15(1):77–87. Available from: https://doi.org/10.1016/j.cpet.2019.08.004

83. Laurens ST, Oyen WJG. Impact of fluorodeoxyglucose PET/computed tomography on the management of patients with colorectal cancer. PET Clin [Internet]. 2015;10(3):345–60. Available from: http://dx.doi.org/10.1016/j.cpet.2015.03.007

84. Cornelis FH, Petre EN, Vakiani E, Klimstra D, Durack JC, Gonen M, et al. Immediate postablation 18 F-FDG injection and corresponding SUV are surrogate biomarkers of local tumour progression after thermal ablation of colorectal carcinoma liver metastases. J Nucl Med. 2018;59(9):1360–5.

85. Cornelis F, Storchios V, Violari E, Sofocleous CT, Schoder H, Durack JC, et al. 18F-FDG PET/CT is an immediate imaging biomarker of treatment success after liver metastasis ablation. J Nucl Med. 2016;57(7):1052–7.

86. Walck C. Hand-book on STATISTICAL DISTRIBUTIONS for experimentalists. Univ Stock. 2007;

87. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. JAMA - J Am Med Assoc. 2016;316(10):1093–103.

88. Büyükkaramikli NC, Rutten-van Mölken MPMH, Severens JL, Al M. TECH-VER: A Verification Checklist to Reduce Errors in Models and Improve Their Credibility. Pharmacoeconomics [Internet]. 2019;37(11):1391–408. Available from: https://doi.org/10.1007/s40273-019-00844-y

89. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty: A report of the ISPOR-SMDM modeling good research practices task force-6. Value Heal. 2012;15(6):835–42.

90. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using discrete event simulation: A report of the ISPOR-SMDM modeling good research practices task force-4. Value Heal [Internet]. 2012;15(6):821–7. Available from: http://dx.doi.org/10.1016/j.jval.2012.04.013

91. Degeling K, Koffijberg H, IJzerman MJ. A systematic review and checklist presenting the main challenges for health economic modeling in personalized medicine: towards implementing patient-level models. Expert Rev Pharmacoeconomics Outcomes Res [Internet]. 2017;17(1):17–25. Available from: https://doi.org/10.1080/14737167.2017.1273110

92. Pitt M, Monks T, Crowe S, Vasilakis C. Systems modelling and simulation in health service design, delivery and decision making. BMJ Qual Saf. 2016;25(1):38–45.

93. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: A report of the ISPOR-SMDM modeling good research practices task force-2. Med Decis Mak. 2012;32(5):678–89.

94. Ucar I, Smeets B, Azcorra A. simmer: Discrete-Event Simulation for R. J Stat Softw. 2019;90(2):1–30.

95. Al B et. Bayesian Cost Effectiveness Analysis with the R package BCEA. Springer, New York. 2017;

96. CBS. Levensverwachting, geslacht, leeftijd (per jaar en periode van vijf jaren) [Internet]. 2021 [cited 2022 Apr 17]. Available from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37360ned/table?dl=4CA57

97. Ouwens MJNM, Mukhopadhyay P, Zhang Y, Huang M, Latimer N, Briggs A. Estimating Lifetime Benefits Associated with Immuno-Oncology Therapies: Challenges and Approaches for Overall Survival Extrapolations. Vol. 37, PharmacoEconomics. 2019. p. 1129–38.

98. Vickers A. An Evaluation of Survival Curve Extrapolation Techniques UsingLong-Term Observational Cancer Data. Med Decis Mak. 2019;39(8):926–38.

99. Bagust A, Beale S. Survival Analysis And Extrapolation Modeling Of TTE Clinical Trial Data For Economic Evaluation. Med Decis Mak. 2013;

100. Davies C, Briggs A, Lorgelly P, Garellick G, Malchau H. The "hazards" of extrapolating survival curves. Med Decis Mak. 2013;33(3):369–80.

101. Jackson C, Stevens J, Ren S, Latimer N, Bojke L, Manca A, et al. Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods. Med Decis Mak. 2017;37(4):377–90.

102. Kearns B, Stevens J, Ren S, Brennan A. How Uncertain is the Survival Extrapolation? A Study of the Impact of Different Parametric Survival Models on Extrapolated Uncertainty About Hazard Functions, Lifetime Mean Survival and Cost Effectiveness. Vol. 38, PharmacoEconomics. 2020. p. 193–204.

103. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: Reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol. 2012;12.

104. Wei Y, Royston P. Reconstructing Time-to-event Data from Published Kaplan–Meier Curves. Stata J Promot Commun Stat Stata. 2017;17(4):786–802.

105. Bijlstra OD, Boreel MME, van Mossel S, Burgmans MC, Kapiteijn EHW, Oprea-Lager DE, et al. The Value of 18F-FDG-PET-CT Imaging in Treatment Evaluation of Colorectal Liver Metastases: A Systematic Review. Diagnostics. 2022;12(3):715.

106. Oostenbrink JB, Koopmanschap MA, Rutten FFH. Standardisation of costs: The Dutch Manual for Costing in Economic Evaluations. Pharmacoeconomics. 2002;20(7):443–54.

107. IJzerman MJ. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 2016.

108. Pomp M, Brouwer W, Rutten F. QALY-tijd - Nieuwe medische technologie, kosteneffectiviteit en richtlijnen. Cent Planbur. 2007;152:1–86.

109. Hakkaart-van Roijen L, Van der Linden N, Bouwmans C, Kanters T, Swan Tan S. Kostenhandleiding : Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. Erasmas Univ Rotterdam. 2015;1–120.

110. LUMC. Passanten prijslijst DBC-zorgproducten en overige zorgproducten jaar 2021 - LUMC - ingangsdatum 1-jul-2021. 2021. p. 1–143.

111. Jeong K, Cairns J. Systematic review of health state utility values for economic evaluation of colorectal cancer. Health Econ Rev [Internet]. 2016;6(1):15–7. Available from: http://dx.doi.org/10.1186/s13561-016-0115-5

112. Wiering B, Oyen WJG, Adang EMM, Van Der Sijp JRM, Roumen RM, De Jong KP, et al. Long-term global quality of life in patients treated for colorectal liver metastases. Br J Surg. 2011;98(4):565–71.

113. Degeling K, Koffijberg H, Franken MD, Koopman M, IJzerman MJ. Comparing Strategies for Modeling Competing Risks in Discrete-Event Simulations: A Simulation Study and Illustration in Colorectal Cancer. Med Decis Mak. 2019;39(1):57–73.

114. Meltzer MI. Introduction to health economics for physicians. Lancet. 2001;358(9286):993–8.

115. Bambha K, Kim WR. Cost-effectiveness analysis and incremental cost-effectiveness ratios: Uses and pitfalls. Eur J Gastroenterol Hepatol. 2004;16(6):519–26.

116. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Fourth edi. Oxford University Press; 2015.

117. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: The role of cost-effectiveness acceptability curves. Health Econ. 2001;10(8):779–87.

118. Claxton K. The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999;18(3):341–64.

119. CONSIDERING AND COMMUNICATING UNCERTAINTY IN HEALTH TECHNOLOGY HTAi Global Policy Forum 2021 Background Paper. HTAi Gloabl Policy Forum. 2021;(February).

120. Fenwick E, Steuten L, Knies S, Ghabri S, Basu A, Murray JF, et al. Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. Value Heal. 2020;23(2):139–50.

121. Rothery C, Strong M, Koffijberg H (Erik), Basu A, Ghabri S, Knies S, et al. Value of Information Analytical Methods: Report 2 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. Value Heal. 2020;23(3):277–86.

122. Claxton, Palmer, Bojke, Sculpher, Philips. Priority setting for research in health care: An application of value of information analysis to glycoprotein IIb/IIIa antagonists in non-ST elevation acute coronary syndrome. [cited 2022 May 25]; Available from: https://doi.org/10.1017/S0266462306051282

123. Degeling K, Vu M, Koffijberg H, Wong HL, Koopman M, Gibbs P, et al. Health Economic Models for Metastatic Colorectal Cancer: A Methodological Review [Internet]. Vol. 38, PharmacoEconomics. Springer International Publishing; 2020. 683–713 p. Available from: https://doi.org/10.1007/s40273-020-00908-4

124. Degeling K, Franken MD, May AM, van Oijen MGH, Koopman M, Punt CJA, et al. Matching the model with the evidence: comparing discrete event simulation and state-transition modeling for time-to-event predictions in a cost-effectiveness analysis of treatment in metastatic colorectal cancer patients. Cancer Epidemiol [Internet]. 2018;57(June):60–7. Available from: https://doi.org/10.1016/j.canep.2018.09.008

125. Degeling K, Wong HL, Koffijberg H, Jalali A, Shapiro J, Kosmider S, et al. Simulating Progression-Free and Overall Survival for First-Line Doublet Chemotherapy With or Without Bevacizumab in Metastatic Colorectal Cancer Patients Based on Real-World Registry Data. Pharmacoeconomics [Internet]. 2020;38(11):1263–75. Available from: https://doi.org/10.1007/s40273-020-00951-1

126. Krijkamp EM, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MGM, Pechlivanoglou P. Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial. Med Decis Mak. 2018;38(3):400–22.

127. Maas M, Beets-Tan R, Gaubert JY, Gomez Munoz F, Habert P, Klompenhouwer LG, et al. Follow-up after radiological intervention in oncology: ECIO-ESOI evidence and consensus-based recommendations for clinical practice. Insights Imaging. 2020;11(1).

128. García-Figueiras R, Baleato-González S, Padhani AR, Luna-Alcalá A, Marhuenda A, Vilanova JC, et al. Advanced imaging techniques in evaluation of colorectal cancer. Radiographics. 2018;38(3):740–65.

129. Viganò L, Lopci E, Costa G, Rodari M, Poretti D, Pedicini V, et al. Positron Emission Tomography-Computed Tomography for Patients with Recurrent Colorectal Liver Metastases: Impact on Restaging and Treatment Planning. Ann Surg Oncol. 2017;24(4):1029–36.

130. Khan K, Athauda A, Aitken K, Cunningham D, Watkins D, Starling N, et al. Survival Outcomes in Asymptomatic Patients With Normal Conventional Imaging but Raised Carcinoembryonic Antigen Levels in Colorectal Cancer Following Positron Emission Tomography-Computed Tomography Imaging. Oncologist. 2016;21(12):1502–8.

131. Monteil J, Le Brun-Ly V, Cachin F, Zasadny X, Seitz JF, Mundler O, et al. Comparison of 18FDG-PET/CT and conventional follow-up methods in colorectal cancer: A randomised prospective study. Dig Liver Dis [Internet]. 2021;53(2):231–7. Available from: https://doi.org/10.1016/j.dld.2020.10.012

132. Sahani D V., Bajwa MA, Andrabi Y, Bajpai S, Cusack JC. Current status of imaging and emerging techniques to evaluate liver metastases from colorectal carcinoma. Ann Surg. 2014;259(5):861–72.

133. Fowler KJ, Linehan DC, Menias CO. Colorectal liver metastases: State of the art imaging. Ann Surg Oncol. 2013;20(4):1185–93.

134. Gallamini A, Zwarthoed C, Borra A. Positron emission tomography (PET) in oncology. Cancers (Basel). 2014;6(4):1821–89.

135. Renzulli M, Clemente A, Ierardi AM, Pettinari I, Tovoli F, Brocchi S, et al. Imaging of colorectal liver metastases: New developments and pending issues. Cancers (Basel). 2020;12(1).

136. Travaini LL, Trifirò G, Ravasi L, Monfardini L, Della Vigna P, Bonomo G, et al. Role of [18F]FDG-PET/CT after radiofrequency ablation of liver metastases: Preliminary results. Eur J Nucl Med Mol Imaging. 2008;35(7):1316–22.

137. Kuehl H, Stattaus J, Hertel S, Hunold P, Kaiser G, Bockisch A, et al. Mid-term Outcome of Positron Emission Tomography/Computed Tomography-assisted Radiofrequency Ablation in Primary and Secondary Liver Tumours - A Single-centre Experience. Clin Oncol. 2008;20(3):234–40.

138. Kuehl H, Antoch G, Stergar H, Veit-Haibach P, Rosenbaum-Krumme S, Vogt F, et al. Comparison of FDG-PET, PET/CT and MRI for follow-up of colorectal liver metastases treated with radiofrequency ablation: Initial results. Eur J Radiol. 2008;67(2):362–71.

139. Veit P, Antoch G, Stergar H, Bockisch A, Forsting M, Kuehl H. Detection of residual tumour after radiofrequency ablation of liver metastasis with dual-modality PET/CT: Initial results. Eur Radiol. 2006;16(1):80–7.

140. Rojas Llimpe FL, Di Fabio F, Ercolani G, Giampalma E, Cappelli A, Serra C, et al. Imaging in resectable colorectal liver metastasis patients with or without preoperative chemotherapy: Results of the PROMETEO-01 study. Br J Cancer. 2014;111(4):667–73.

141. Gade M, Kubik M, Fisker R V., Thorlacius-Ussing O, Petersen LJ. Diagnostic value of 18F-FDG PET/CT as first choice in the detection of recurrent colorectal cancer due to rising CEA. Cancer Imaging [Internet]. 2015;15(1):1–8. Available from: http://dx.doi.org/10.1186/s40644-015-0048-y

142. Sahin DA, Agcaoglu O, Chretien C, Siperstein A, Berber E. The utility of PET/CT in the management of patients with colorectal liver metastases undergoing laparascopic radiofrequency thermal ablation. Ann Surg Oncol. 2012;19(3):850–5.

143. Nielsen K, Van Tilborg AAJM, Scheffer HJ, Meijerink MR, De Lange-De Klerk ESM, Meijer S, et al. PET-CT after radiofrequency ablation of colorectal liver metastases:Suggestions for timing and image interpretation. Eur J Radiol [Internet]. 2013;82(12):2169–75. Available from: http://dx.doi.org/10.1016/j.ejrad.2013.08.024

144. Rothman KJ, Greenland S, Associate TLL. Modern Epidemiology, 3rd Edition. Hastings Cent Rep [Internet]. 2014 [cited 2022 May 10];44 Suppl 2:insidebackcover. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24644503

145. Adarkwah CC, van Gils PF, Hiligsmann M, Evers SMAA. Risk of bias in model-based economic evaluations: the ECOBIAS checklist. Expert Rev Pharmacoeconomics Outcomes Res [Internet]. 2016;16(4):513–23. Available from: http://dx.doi.org/10.1586/14737167.2015.1103185

146. Van Der Stok EP, Spaander MCW, Grünhagen DJ, Verhoef C, Kuipers EJ. Surveillance after curative treatment for colorectal cancer. Nat Rev Clin Oncol. 2017;14(5):297–315.

147. Okamura R, Hida K, Nishizaki D, Sugihara K, Sakai Y. Proposal of a stage-specific surveillance strategy for colorectal cancer patients: A retrospective analysis of Japanese large cohort. Eur J Surg Oncol [Internet]. 2018;44(4):449–55. Available from: https://doi.org/10.1016/j.ejso.2018.01.080

148. Sensible and sustainable care. Counc Public Heal Heal Care. 2006;

149. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, et al. Modelling in economic evaluation: An unavoidable fact of life. Health Econ. 1997;6(3):217–27.

150. Baio G. survHE: Survival Analysis for Health Economic Evaluation and Cost-Effectiveness Modeling. J Stat Softw. 2020;95(14):1–47.

Appendix A

A probability of non-disease-related death was derived from available Dutch demographic data. This all-cause probability of death was assigned to each patient based on the age of the patient. Table 21 visualises the Dutch averaged age-dependent survival times for the period 2016-2020.

Table 21 Averaged Dutch age-dependent and non-disease-related life expectancy (period 2016-2020)

Age	Mortality rate	Cumulative survival rate	Life expectancy
0	0,003218	0,996782	81,696
1	0,000464	0,996319	81,46
2	0,000184	0,996136	80,496
3	0,000112	0,996025	79,51
4	0,000104	0,995921	78,52
5	0,00007	0,995851	77,53
6	0,000096	0,995756	76,53
7	0,000056	0,9957	75,54
8	0,000064	0,995636	74,544
9	0,00006	0,995576	73,55
10	0,000058	0,995519	72,556
11	0,000058	0,995461	71,556
12	0,000078	0,995383	70,564
13	0,000078	0,995306	69,566
14	0,000108	0,995198	68,572
15	0,00013	0,995069	67,58
16	0,000148	0,994922	66,59
17	0,000166	0,994756	65,6
18	0,000218	0,99454	64,612
19	0,000254	0,994287	63,624
20	0,000292	0,993997	62,638
21	0,000258	0,99374	61,658
22	0,000284	0,993458	60,674
23	0,000292	0,993168	59,69
24	0,000288	0,992882	58,706
25	0,000292	0,992592	57,722
26	0,000286	0,992308	56,742
27	0,000342	0,991969	55,758
28	0,000346	0,991625	54,774
29	0,00036	0,991268	53,794
30	0,00038	0,990892	52,812
31	0,000384	0,990511	51,832
32	0,000428	0,990087	50,854
33	0,00049	0,989602	49,876
34	0,00045	0,989157	48,9
35	0,000538	0,988625	47,922

36	0,000542	0,988089	46,946
37	0,000572	0,987524	45,972
38	0,000652	0,98688	45
39	0,00068	0,986209	44,026
40	0,000712	0,985507	43,056
41	0,00082	0,984698	42,086
42	0,000894	0,983818	41,12
43	0,001008	0,982826	40,158
44	0,001128	0,981718	39,198
45	0,001224	0,980516	38,24
46	0,001322	0,97922	37,288
47	0,001492	0,977759	36,334
48	0,001664	0,976132	35,39
49	0,001784	0,974391	34,448
50	0,001922	0,972518	33,51
51	0,002266	0,970314	32,572
52	0,002536	0,967853	31,646
53	0,00276	0,965182	30,724
54	0,00304	0,962248	29,806
55	0,003412	0,958965	28,896
56	0,003686	0,95543	27,992
57	0,004156	0,951459	27,096
58	0,004562	0,947119	26,206
59	0,00511	0,942279	25,326
60	0,005664	0,936942	24,452
61	0,006308	0,931032	23,59
62	0,007004	0,924511	22,738
63	0,007618	0,917468	21,89
64	0,008454	0,909711	21,054
65	0,009124	0,901411	20,23
66	0,010092	0,892314	19,412
67	0,010906	0,882583	18,606
68	0,011992	0,871999	17,806
69	0,013132	0,860548	17,014
70	0,014454	0,848109	16,234
71	0,016008	0,834533	15,466
72	0,017804	0,819675	14,71
73	0,01963	0,803584	13,966
74	0,021796	0,78607	13,234
75	0,024324	0,766949	12,518
76	0,027284	0,746024	11,822
77	0,030046	0,723609	11,138
78	0,033482	0,699381	10,466
79	0,03787	0,672895	9,81
80	0,043184	0,643837	9,176

81	0,048262	0,612764	8,568
82	0,055068	0,57902	7,974
83	0,062608	0,542769	7,412
84	0,070844	0,504317	6,874
85	0,080656	0,463641	6,36
86	0,092662	0,420679	5,874
87	0,105016	0,376501	5,424
88	0,119386	0,331552	4,998
89	0,134266	0,287036	4,606
90	0,152346	0,243307	4,246
91	0,168792	0,202239	3,916
92	0,187974	0,164223	3,61
93	0,209098	0,129884	3,33
94	0,231226	0,099852	3,08
95	0,25312	0,074577	2,854
96	0,27637	0,053966	2,646
97	0,303366	0,037595	2,472
98	0,318636	0,025616	2,33
99	0,373104	0,016058	2,184

Appendix B

The Kaplan-Meier (KM) estimator is a non-parametric statistic used to estimate the survival function from lifetime data. To enhance the quality of secondary data analyses, we applied a method which derives from published KM survival curves a close approximation to the original individual patient-level time-to-event data from which they were generated. The algorithm of *Guyot et al*. maps from digitized curves back to KM data by finding numerical solutions to the inverted KM equations using the information on the number of events and the numbers at risk.

The algorithm of *Guyot et al.* is incorporated in *R* version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria; http://www.R-project.org/), package *survHE* version 1.1.2 (150).

Notice that we checked and justified which distribution was best by visual checks (e.g., Q-Q plots, P-P plots, histograms and density plots) and statistical checks (e.g., AIC, BIC). Final results are given in Figures 8 and 9. For more detailed information, see the accompanied r-file named 'Data Analysis'.

In Figure 8, the time points, when reaching the median values of both downstream treatment strategies[1], clearly vary in favour of rLAT. The reason is that patients receiving LAT have, a priori, better survival probabilities (e.g., these are patients with, on average, less tumour spread and a better clinical condition).

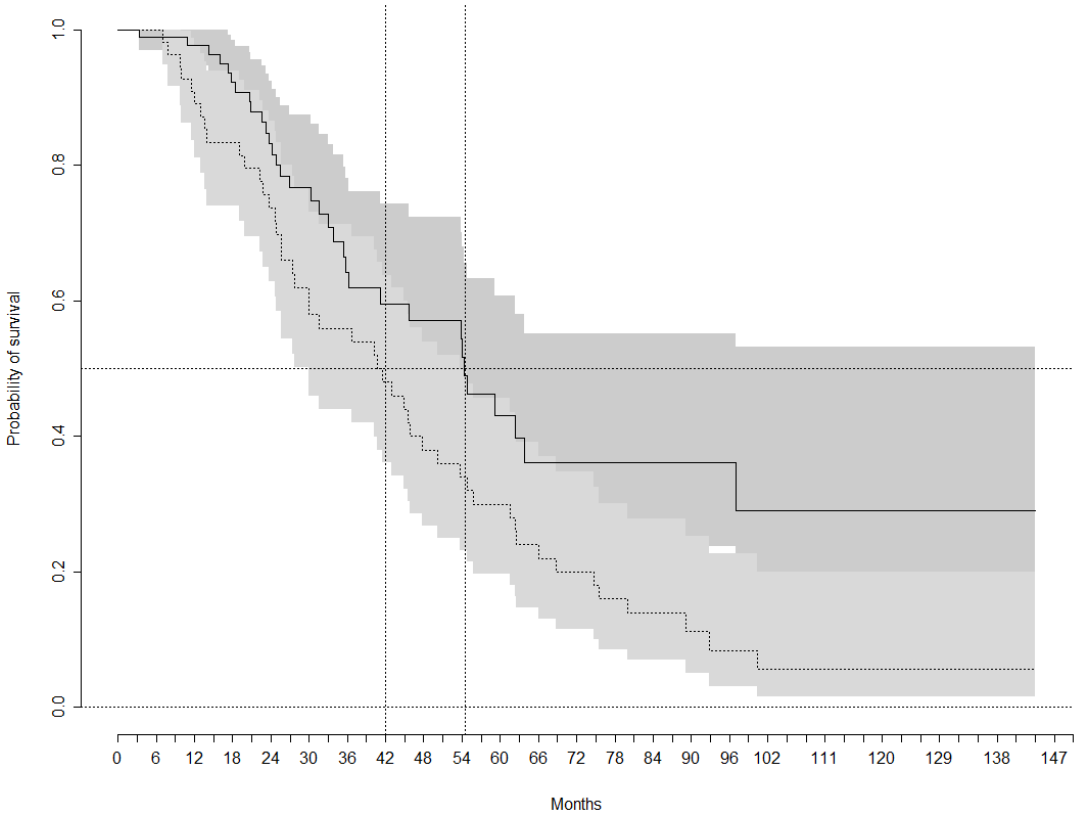

Figure 8 Visualised overall survival times for downstream rLAT (solid line) and systemic therapy (dashed line)

[1] The median value was assumed to be 0.5 because each patient starts in perfect health (value=1) and will eventually not survive (value=0).

Local/residual recurrence in the ablation or resection zone following TA therapy indicates the strength of the provided intervention. Intrahepatic non-local and extrahepatic distant metastases indicate the degree of aggressiveness of colorectal liver disease. Early tumour detection ensures that we minimalize extrahepatic tumour spread. Also, early detection ensures that tumours are relatively small and can be treated with minimally invasive interventions with minimal risks of adverse events.

Figure 9, below, underlines these considerations. It shows the PFS rates following TA therapy. It is expected that residual/local recurrence is, on average, detected earlier (in time) but less often (in number) compared to intrahepatic non-local and extrahepatic distant CRLM respectively.

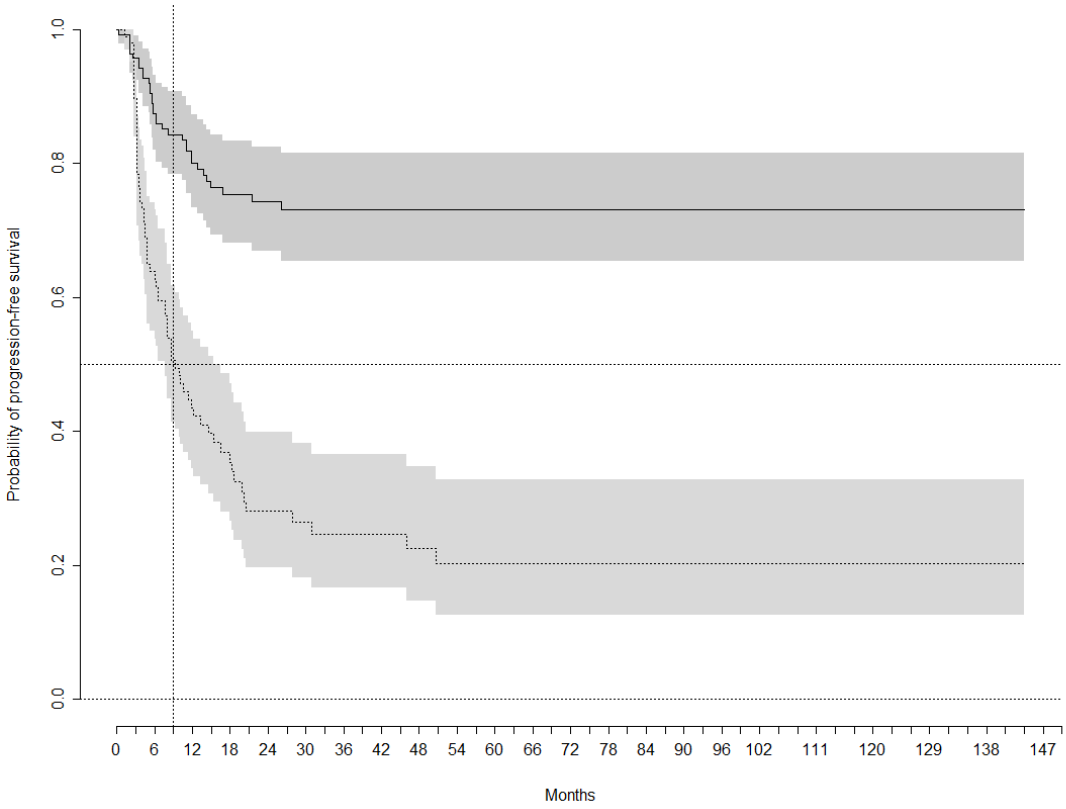

Figure 9 Visualised PFS rates after TA therapy. The solid line shows the PFS rates for residual/local recurrence. The dashed line shows the PFS rates for intrahepatic non-local and extrahepatic distant recurrence respectively

Appendix C

The applied literature-based evidence on HR-QoLs was calculated from the EQ-5D index scores and can be found in Tables 22-25 (112). The utility tables reflect HR-QoL values for groups of patients (that is the average utilities). The tables are categorized by the defined disease states **(1)** and the downstream treatment strategy received **(2)**. Tables 22 and 23 reflect HR-QoL values in case of developing no recurrence, respectively, developing recurrence. Tables 24 and 25 reflect HR-QoL values in the case of rLAT, respectively, systemic therapy received. Note that all HR-QoL values were truncated for a maximum upper value and a minimum lower value resulting in a restricted predefined range of [-0,59;0.92] (112).

Table 22 Reflecting the disease state: developing no recurrence (112)

Time after TA therapy received	Mean HR-QoL value	SEM	No.	SD
0 (=baseline value)	0,8202	0,012209	117	0,132057
3 weeks	0,573807	0,027747	103	0,281601
6 weeks	0,72586	0,023307	97	0,229552
3 months	0,798002	0,018313	91	0,174695
6 months	0,830189	0,023862	70	0,199647
9 months	0,810211	0,024972	60	0,193434
12 months	0,833518	0,022752	54	0,167196
15 months	0,82242	0,022752	43	0,149198
18 months	0,781354	0,039401	43	0,258367
21 months	0,804661	0,036071	41	0,230967
24 months	0,804661	0,036071	38	0,222357
27 months	0,833518	0,024417	38	0,150518
30 months	0,809101	0,027747	35	0,164153
33 months	0,780244	0,04717	30	0,25836
36 months	0,758047	0,045505	34	0,265337

Abbreviations: SEM=standard error of the mean, No.=number of patients, SD=standard deviation

Table 23 Reflecting the disease state: developing local/residual, non-local and/or distant recurrence (112)

Time after TA therapy received	Mean HR-QoL value	SEM	No.	SD
0 (=baseline value)	0,8202	0,012209	117	0,132057
3 weeks	0,573807	0,027747	103	0,281601
6 weeks	0,72586	0,023307	97	0,229552
3 months	0,784205	0,031146	16	0,124583
6 months	0,660734	0,058954	34	0,34376
9 months	0,720801	0,047831	37	0,290944
12 months	0,806452	0,021691	41	0,138889
15 months	0,709677	0,031146	45	0,208932
18 months	0,751947	0,027253	44	0,180773
21 months	0,756396	0,025584	44	0,169705
24 months	0,760845	0,036151	43	0,23706

27 months	0,765295	0,03337	39	0,208398
30 months	0,746385	0,035595	37	0,216517
33 months	0,72525	0,039488	33	0,226843
36 months	0,570634	0,064516	36	0,387097

Abbreviations: SEM=standard error of the mean, No.=number of patients, SD=standard deviation

Table 24 Reflecting the downstream treatment: rLAT received (112)

Time after LAT received	Mean HR-QoL value	SEM	No.	SD
3 months	0,800441	0,03473	12	0,120308
6 months	0,735391	0,067255	21	0,3082
9 months	0,803749	0,019846	19	0,086505
12 months	0,853363	0,015987	21	0,073261
15 months	0,768467	0,021499	20	0,096148
18 months	0,848953	0,017089	20	0,076426
21 months	0,841235	0,023153	19	0,100923
24 months	0,833517	0,024256	17	0,100009
27 months	0,818082	0,059537	16	0,238148
30 months	0,850055	0,025358	13	0,091431
33 months	0,767365	0,07387	10	0,233597
36 months	0,733186	0,07387	15	0,286097

Abbreviations: SEM=standard error of the mean, No.=number of patients, SD=standard deviation

Table 25 Reflecting the downstream treatment: systemic therapy received (112)

Time after systemic therapy received	Mean HR-QoL value	SEM	No.	SD
3 months	0,54415	0,102097	13	0,368116
6 months	0,54415	0,102097	13	0,368116
9 months	0,636865	0,088852	18	0,376968
12 months	0,754967	0,038079	20	0,170297
15 months	0,666667	0,050773	25	0,253863
18 months	0,675497	0,048565	24	0,23792
21 months	0,696468	0,04415	25	0,220751
24 months	0,713024	0,056291	26	0,287031
27 months	0,737307	0,039183	23	0,187916
30 months	0,699779	0,05298	24	0,259549
33 months	0,713024	0,048565	23	0,23291
36 months	0,453642	0,094923	21	0,434991

Abbreviations: SEM=standard error of the mean, No.=number of patients, SD=standard deviation

Appendix D

To determine the number of samples (=patients) to simulate, we provided a base case analysis (BCA). In this analysis, we did not incorporate parameter uncertainty, however, we incorporated stochastic uncertainty. Idea was that we wanted to determine the number of patients to simulate, which should be a sufficiently large number to remove the impact of patient-level variation in our cost-effectiveness outcomes. Therefore, we needed to get insights into the distribution of the data of the outcomes of our interest (e.g., the amount of QALYs gained and/or the costs made for the intervention group and comparator group respectively). An appropriate method for displaying this was by providing boxplots, which show us if our data is symmetric, tightly grouped and skewed while increasing the number of simulated patients. It also shows us outliers (point estimates ≥ 1.5*Interquartile Range + the 75th percentile of our data). On the next pages, we show such boxplots for the costs made, the amount of LYGs and the amount of QALYs gained when a patient walks through the CRLM care pathway, given, a fixed number of 100 runs provided (Figures 10-15).

Table 26 shows the calculated ICER values for BCA while increasing the number of simulated patients (given a fixed number of 100 runs). Here, it was essential to assess their stability while only accepting a decent computation time. Finally, it was decided to simulate a minimum of 10,000 patients per run.

Table 26 ICER estimates while increasing the number of simulated patients

Number of patients to simulate	**ICER *(€/QALY)***
1.000 (x100 runs)	€18,256.17
5.000 (x100 runs)	€15,999.06
10.000 (x100 runs)	€29,931.51
20.000 (x100 runs)	€17,130.20
50.000 (x100 runs)	€15,071.39

Note that we did not run a certain amount of Monte Carlo samples just because the error in the incremental costs/effects then is sufficiently small. Such thresholds of statistical significance are arbitrary and completely ignore the magnitude of making the (wrong) decision (118)**.**

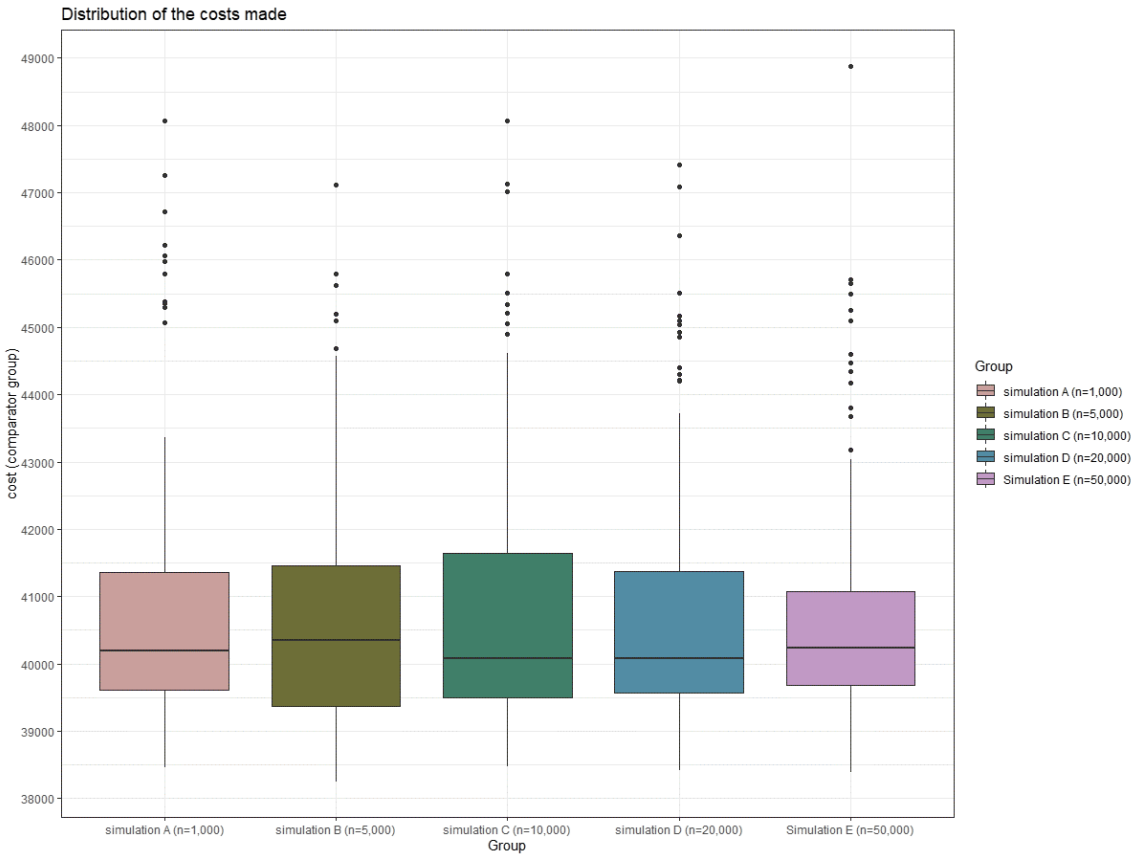

*Figure 10 Distribution of the costs made while increasing the number of simulated patients (**comparator group**)*

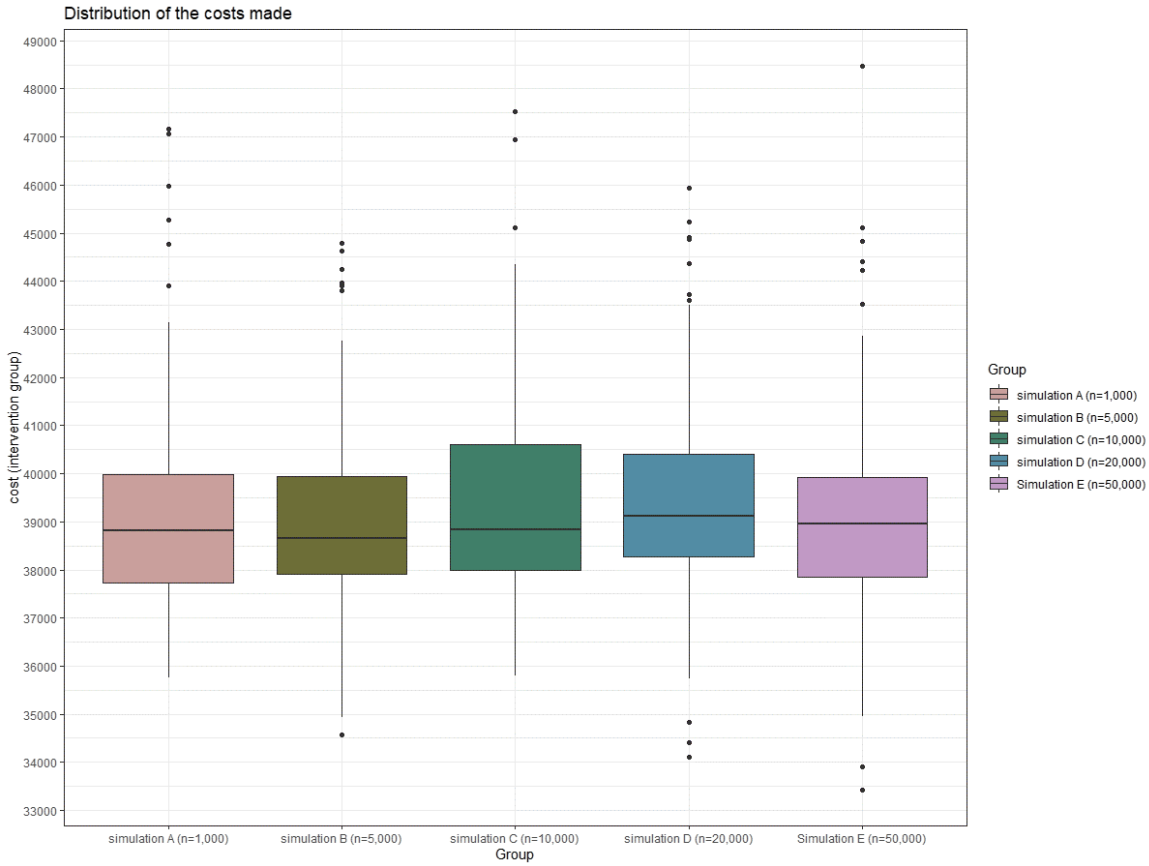

*Figure 11 Distribution of the costs made while increasing the number of simulated patients (**intervention group**)*

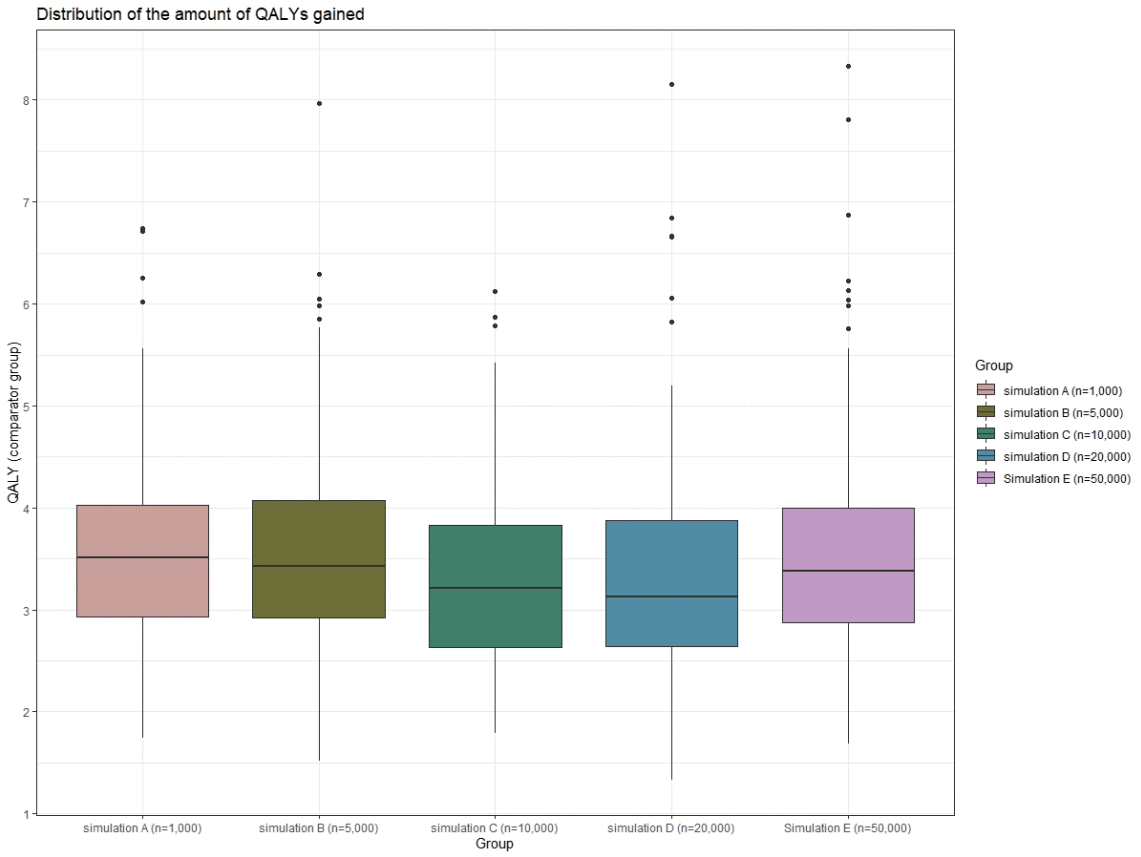

*Figure 12 Distribution of the amount of QALYs gained while increasing the simulated patients (**comparator group**)*

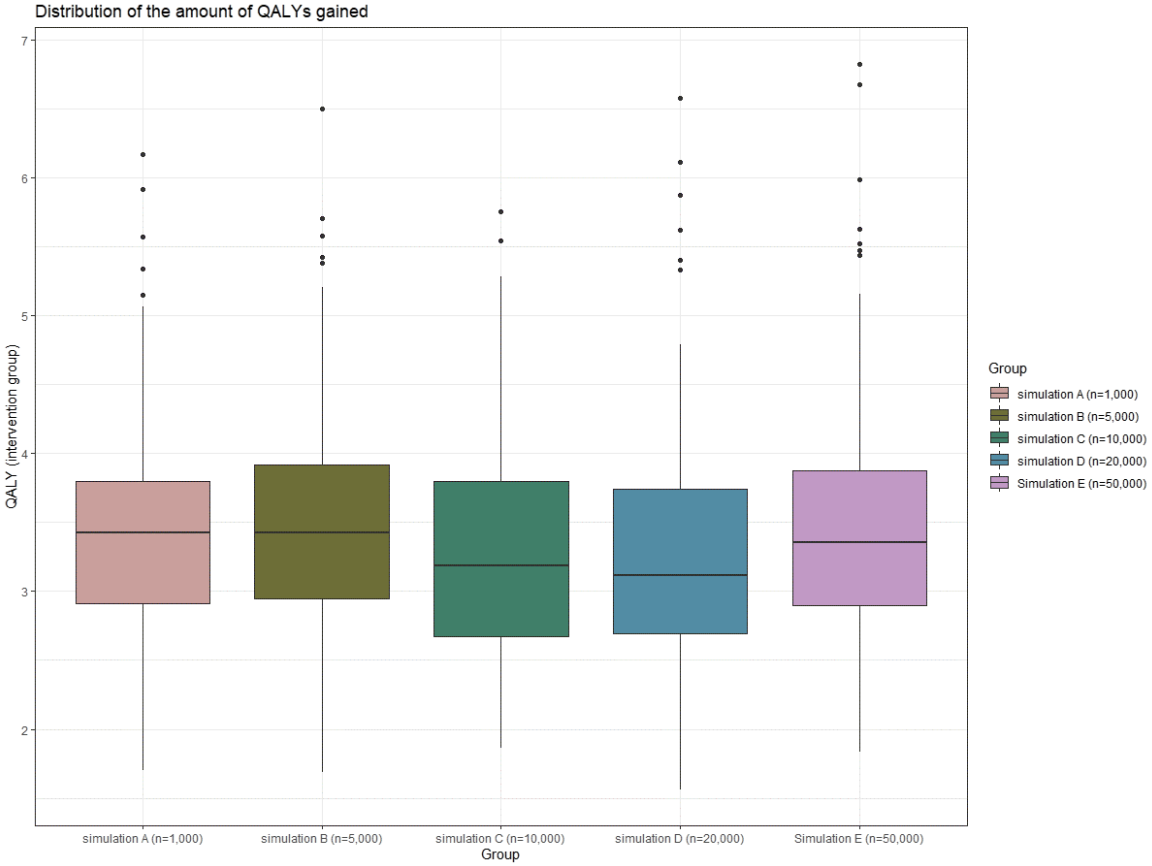

*Figure 13 Distribution of the amount of QALYs gained while increasing the simulated patients (**intervention group**)*

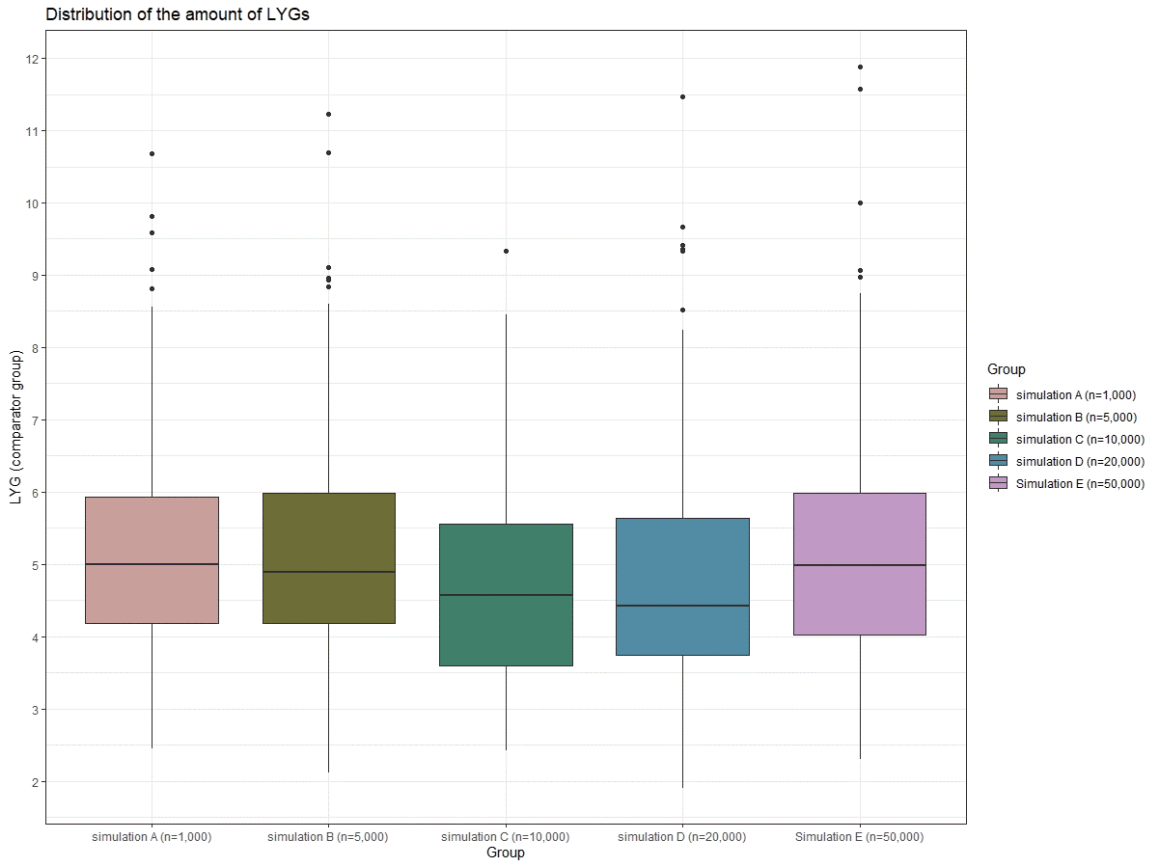

*Figure 14 Distribution of the amount of LYGs while increasing the number of simulated patients (**comparator group**)*

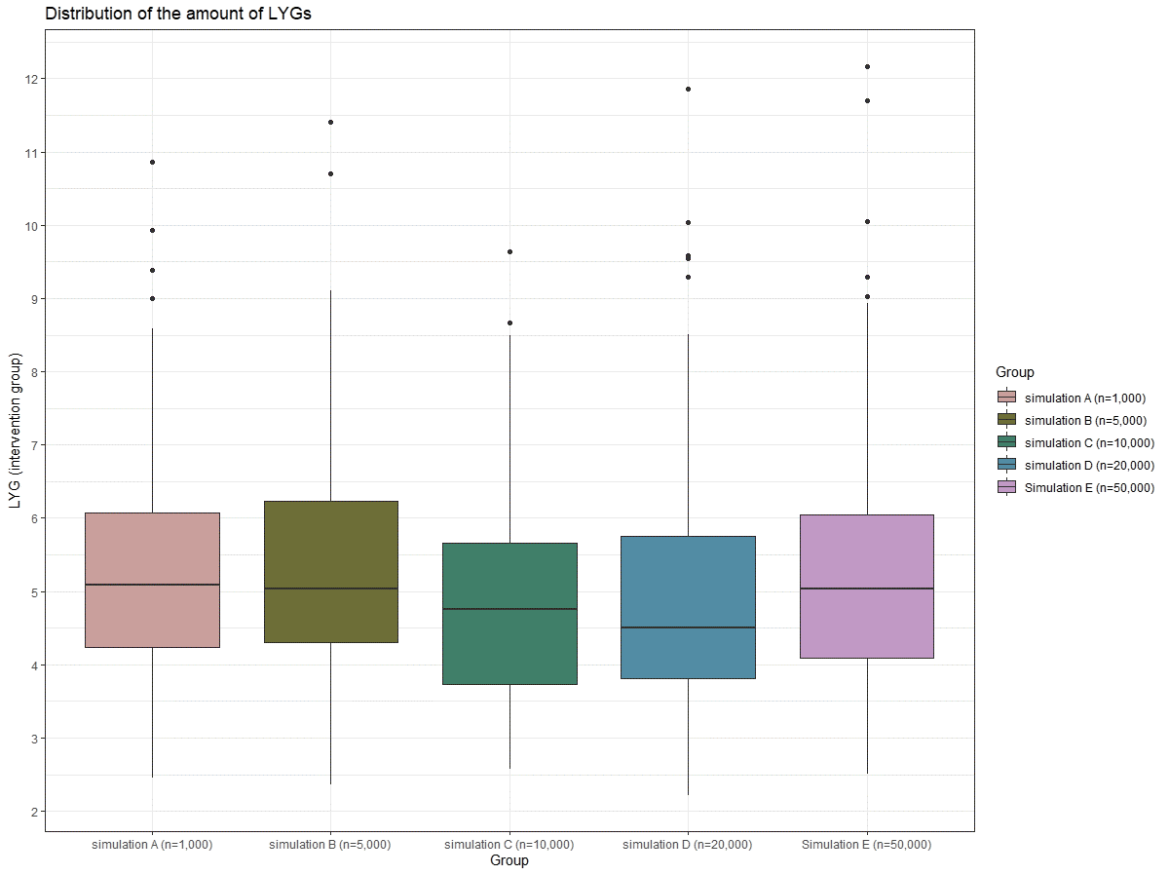

*Figure 15 Distribution of the amount of LYGs while increasing the number of simulated patients (**intervention group**)*