**Textbook outcome after trans-arterial embolization for hepatocellular carcinoma**

Cristina Mosconi1, Joanne O’Rourke2, Roman Kloeckner3, Lukas Sturm4,5, Rita Golfieri1, Ciro Celsa6, Waleed Fateen7, Bruno C. Odisio8, Enrico Matteo Garanzini9, Markus Peck10, Alberto Borghi11, Yuk Ting Ma2, Fabian Stoehr3, Dominik Bettinger4,5, Paolo Giuffrida6, Guru Aithal7, Yuan-Mao Lin8, Carlo Spreafico9, Emanuela Giampalma12, Philip Johnson13 and Alessandro Cucchetti14

Collaborators: Niccolò Brandi1, Alberta Cappelli1,Francesco Giuseppe Foschi11, Giuseppe Cabibbo6, Leonardo Solaini14, Adina Olaru7, Simon Travis7, Aloysious Aravinthan7

1 Department of Diagnostic Medicine and Prevention, Radiology Unit, Azienda Ospedaliero-Universitaria di Bologna, IRCCS, Sant'Orsola-Malpighi Hospital, Bologna, Italy

2 Centre for Liver and Gastrointestinal Research, Institute of Immunology and Immunotherapy, Medical School, University of Birmingham, Birmingham, United Kingdom.

3 Department of Diagnostic and Interventional Radiology, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany

4 Department of Medicine II, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Hugstetter Str. 55, D-79106 Freiburg, German

5 Berta-Ottenstein-Programme, Faculty of Medicine, University of Freiburg

6 Section of Gastroenterology & Hepatology, Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties, PROMISE, University of Palermo, Italy

7 NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and The University of Nottingham, Nottingham, UK

8 Department of Interventional Radiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

9 Department of Radiology, Fondazione IRCCS Istituto Nazionale Tumori di Milano, Milan, Italy.

10 Department of Internal Medicine and Gastroenterology, Klinikum Klagenfurt am Wörthersee, Klagenfurt, Austria.

11 Department of Internal Medicine, Infermi Hospital, Area Vasta Romagna, Faenza, Italy.

12 Radiology Unit, Maurizio Bufalini Hospital, Cesena, Italy.

13 Department of Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, UK

14 Department of Medical and Surgical Sciences, Alma Mater Studiorum – University of Bologna, Bologna, Italy; Morgagni – Pierantoni Hospital, Area Vasta Romagna, Forlì, Italy.

**Corresponding author:** Alessandro Cucchetti, MD, Department of Medical and Surgical Sciences – DIMEC; Alma Mater Studiorum - University of Bologna, Via Albertoni 15; 40138 Bologna; ITALY.

**e-mail:** **aleqko@libero.it**

**Running head**: Textbook Outcome after TACE for HCC

**Number of tables and figures:** 4 + 2

**Conflict of interest:** none to declare

**Financial support:** no grant support was used for the present study

**Authors’ contribution:** C.Mosconi had the original idea; A.Cucchetti planned and coordinated the multicenter data collection; J.O’Rourke, R.Kloeckner, L.Sturm, R.Golfieri1, C.Celsa, W. Fateen, B. Odisio, E. Garanzini, M.Peck, A.Borghi, Y.T.Ma, F.Stoehr, D.Bettinger, P.Giuffrida, G.Aithal, Y.Lin, C.Spreafico and Emanuela Giampalma collected clinical data and/or performed the procedures; A. Cucchetti performed the analyses, C.Mosconi and A.Cucchetti wrote the manuscript. P.Johnson provided critical review for important intellectual contents.

**Keywords:** hepatocellular carcinoma; trans-arterial chemoembolization; survival; complications; morbidity; textbook outcome; mRECIST

**ABSTRACT**

**Background & Aims:** quality indicators for interventional procedures must include both safety and efficacy measures. The Textbook Outcome (TO) is inclusive of both but never provided for trans-arterial chemoembolization (TACE) for hepatocellular carcinoma (HCC).

**Methods:** Data on naïve HCC patients treated with TACE were collected from 10 centers at different volumes. TOwas defined as: “*no post-TACE complications, no prolonged hospital stay as defined as a postoperative stay ≤75th percentile of the median values from the total cohort, no readmission or mortality within 30 days and the achievement of an objective response (OR) at the post-TACE imaging."*. Pooled estimates were calculated to account for hospitals’ effect and risk-adjustment applied to handle for the diversity of patients in each center.

**Results:** 1124 patients (2014 – 2018) fulfilling specific inclusion criteria were enrolled. Baseline clinical features showed moderate to considerable heterogeneity (I2) across centers. Absence of TACE-related mortality was 97.6% (95%C.I.: 96.4–98.4; I2: 0.0%), no readmission was observed in 94.9% of patients (95%C.I. 91.6–97.0; I2: 68.6%), no prolonged length of stay in 71.8% (95%C.I.: 56.9–83.1; I2: 95.0%), OR of the target lesion was diagnosed in 68.5% (95%C.I.: 55.6–79.0; I2: 93.6%) and no complication of any grade was observed in 50.0% (95%C.I.: 36.1–63.9; I2: 94.4%). TO was finally achieved in 29.0% of patients (95%C.I.: 20.4–39.5; I2: 90.9%). Median overall survival of patients achieving TO was 35.4 months and that of patients not achieving TO was 24.1 months (HR: 0.57; 95%C.I.: 0.46-0.70). Risk-adjustment showed that TO was similar across involved centers, thus, regardless of hospital volume.

**Conclusions:** TO is a valuable quality indicator after TACE for HCC and it was related with survival, providing early information about long-term survival. Risk-adjustment showed that this outcome was similar across different hospitals.

**LAY SUMMARY:** Textbook Outcome is a composite indicator of safety and efficacy of a specific procedure. The present study assessed this endpoint in patients with hepatocellular carcinoma submitted to trans-arterial chemoembolization from a multi-institutional collaboration of 10 centers. Predictive factors for TO achievement and risk-adjustment through participating centers were provided. TO achievement finally improved patient survival.

**INTRODUCTION**

The provision of good clinical practice is the primary goal of any health system. Any treatment must consider both the expected efficacy and the risks arising from it. A precise evaluation of these two aspects allows to optimize the selection of patients to be treated and allows a qualitative evaluation of the result obtained. To accomplish to these tasks, Textbook Outcome (TO) has been proposed as clinical indicator for measuring health care quality *[Eur J Surg Oncol 2013;39:156–63; BMJ Open. 2018 Mar 1;8(2):e019405]*.

Textbook Outcome is a composite measure consisting of both efficacy and safety aspects within a short-term period. This is particularly suitable and useful for interventional procedures, where early indicators of efficacy can be negative margins in oncologic surgery *[Surgery. 2020 Apr 14:S0039-6060(20)30109-4; JAMA Surg. 2019 Jun 1;154(6):e190571; Ann Surg Oncol. 2020 May 9. doi: 10.1245/s10434-020-08548-w; Br J Surg 2017;104:742–50.]* or the clearance from gallstone disease obtained in ERCP *[BMJ Open. 2018 Mar 1;8(2):e019405]* and where early indicators of safety can be defined through adverse events occurrence, length of hospital stay, post-procedural complications and the eventual need for readmission. Once efficacy and safety indicators are defined, the TO is realized for those patients fulfilling them.

Hepatocellular carcinoma (HCC) is probably the tumor that best lends itself to various surgical and non-surgical procedures. For hepatic resection, TO was already investigated in the past providing that this measure is applicable in this field and providing reference parameters for a qualitative evaluation of the achievable outcome *[Surgery. 2020 Apr 14:S0039-6060(20)30109-4; JAMA Surg. 2019 Jun 1;154(6):e190571; Ann Surg Oncol. 2020 May 9. doi: 10.1245/s10434-020-08548-w]*. Unfortunately, surgery represents one of the least applicable treatments when faced with a diagnosis of HCC, whereas the most frequent adopted therapy is represented by trans-arterial chemoembolization (TACE) *[Liver Int. 2017 Feb;37(2):259-270; J Hepatol. 2017 Jul;67(1):173-183.]*. Nevertheless, such as clinical indicator was never provided for this procedure.

The aim of the present study was thus to assess the likelihood of TO achievement which should be useful for both determine quality of cares, as well as to answer to patient’ needs to know that the *“treatment will be fine”*.

**METHODS**

**Study enrollment**

 A multi-institutional collaboration was planned to involve 10 centers (9 European and one in United States). All participating centers had specific expertise in the management of HCC and the practice of TACE *[Hepatology. 2020 Jul; 72(1): 198–212.]*. During investigators meetings inclusion criteria, necessary data collection and clinical definitions were established. The study protocol and the data collection conformed to the ethical guidelines of the 1975 Declaration of Helsinki and personal data were processed lawfully, fairly and in a transparent manner in relation to the data subject (Regulation (EU) 2016/679 of the European Parliament).

**Inclusion criteria**

 The study planned to enroll patients treated with TACE between January 2014 and December 2018 to ensure a potential follow-up of at least 24 months for more than the 75% of the study population. When defining inclusion criteria, we focused them on patients in which TACE had a common and shared clinical indication. Consequently, patients with portal vein thrombosis, wide‐spread cancer involving more of than 50% of the liver, extrahepatic disease, hepatic encephalopathy and/or refractory ascites were excluded. Regarding liver function, patients with advanced liver impairment were excluded, as measured by means of Child – Pugh score of >B7. Additionally, only patients with Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1 were retained. Finally, eligible patients were those who received TACE as a first-line therapy after a newly diagnosed HCC (naïve patients). The study was per-patient and not per-procedure, thus each individual participant data accounted for the first TACE procedure. Patients submitted to trans-arterial bland embolization (TAE) were not included.

**TACE procedures**

Of the 10 participating centers, 4 declared to perform >200 TACE procedures/year, 4 declared 100-200 procedures/year and 2 declared <100 procedures/year. The adoption of conventional TACE (cTACE) or drug-eluting beads TACE (DEB-TACE) was a discretion of handling centers but in all cases the primary aim was to obtain a super-selective or at least selective cannulation to the HCC feeding artery. The technique of cTACE and DEB-TACE have been already extensively previously described *[Br J Cancer. 2014;111(2):255-264; Cardiovasc Intervent Radiol. 2010;33(1):41-52]*. Different centers also adopted different anesthetic management, being midazolam and/or fentanyl adopted for conscious sedation when clinically indicated, and in relation to the common clinical practice of each center involved. Some participating center in more recent years, cone-beam computed tomography (CBCT) or CT was adopted during TACE for the intraprocedural planning and/or monitoring of the embolized area [Jpn J Radiol. 2011 Jul;29(6):371-7. ]

**Textbook outcome measures**

The following adverse events were recorded within 30 days from TACE: post-embolization syndrome, acute cholecystitis, acute pancreatitis, liver abscess, acute kidney injury, inguinal hematoma, liver dysfunction, length of in-hospital stay and eventual readmission. The grade of each of the adverse events (AE) was based on the Common Terminology Criteria for Adverse Events (CTCAE) v5.0 and Dindo-Clavien suggestions *[Ann Surg. 2004 Aug; 240(2): 205–213].* Details on AE definitions are reported in the *Appendix section*. Post-TACE mortality was defined as death during the initial hospital stay and/or within 30 days after the day of the procedure.

 The most suitable short-term efficacy measure was represented by the radiological assessment after TACE performed at 4-8 weeks. Radiological response to TACE was assessed through mRECIST criteria *[J Hepatol. 2020 Feb;72(2):288-306]*, since this assessment already provided good prognostic information on subsequent long-term survival *[Hepatology. 2019 Nov 7. doi: 10.1002/hep.31022]*. Even though mRECIST assessment can be affected by substantial interobserver variation *[J Hepatol. 2020 Mar 19:S0168-8278(20)30177-X. doi: 10.1016/j.jhep.2020.03.018.]*, the centralized review of relevant scans would not be a solution in the routine clinical practice, thus, we made the pragmatic decision that mRECIST classification, as assessed by the local investigator, would be used in the present study *[Hepatology. 2020 Jul; 72(1): 198–212.]*.

Finally, safety and efficacy items were pooled in the following TO definition: “*no post-TACE complications, no prolonged hospital stay as defined as a postoperative stay ≤75th percentile of the median values from the total cohort, no readmission or mortality within 30 days and the achievement of an objective response (OR = CR + PR) at the post-TACE imaging."* . Notably, considering that TACE is not a potentially curative therapy the main aim should be the control of the disease, thus, OR well accomplish this task.

**Statistical analysis**

Since data derived from a multi-institutional experience, we first investigated heterogeneity among clinical variables potentially able to affect the outcome/s of interest applying a random-effect model *[Control Clin Trials. 1986 Sep;7(3):177-88]*. Heterogeneity (I2) was properly calculated and interpreted as follows <25% = low heterogeneity; 25%-50% = medium, 51%-75% = substantial and >75% = considerable *[Stat Med. 2002;21:1539–58]*. Then, the potential impact of each clinical variable on TO was investigated through multilevel mixed-effects logistic regression (risk-adjustment). This methodology allows to handle for the diversity of patients in each center involved, providing the prediction of what each center's outcome would have been for a standard patient or population, removing the predictable effects of differences across centers *[Stat Med. 1997 Dec 15;16(23):2645-64]*. Once this adjustment is performed, residual differences in outcomes are thought to be related to provider quality. By this way, a risk-standardized hospital TO was finally calculated for each center involved. Finally, the impact of TO achievement on overall survival was tested considering that the final TO achievement was based on radiological response assessment, thus, as a time-dependent covariate. This was accomplished through Cox regression analysis, using the number of involved centers as strata, to provide the most robust variance estimation. Survival was computed from the day of TACE until the last follow-up or death. Transplanted patients were censored the day prior of transplant. No a-priori level of significance was set in the present analyses and potential prognostic factors were considered for multivariable regressions when their confidence intervals (CI) did not include the 1. Collinearity was checked through Variance Inflation Factor (VIF) evaluation.

 Analyses were performed using R-Project 3.2.5 (R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>) and STATA (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.)

**RESULTS**

The study population consisted of 1124 patients treated for HCC as first line therapy. Most of clinical features collected showed considerable heterogeneity (I2>75%) within participating centers, even if weighted values were not dissimilar to the unweighted ones (*Table 1*). Detailed clinical features for each participating center are provided in the *Supplementary Table 1*.

After TACE, 21 patients died and 594 had at least one adverse event (*Table 2*). Detailed data for each participating center are provided in the *Supplementary Table 2*. The weighted mortality was 2.4% and the morbidity was 50.0%. Post-embolization syndrome (PES) of grade 1 or 2, was the most frequent complication, being observed in a weighted proportion of 24.7% of patients, followed by liver dysfunction of grade 1 or 2 which was observed in 13.2% of patients. Both showed considerable heterogeneity across centers (I2>75%). The median values of post-TACE length of stay (LOS) ranged between 1 to 3 days among participating centers, with the 75th percentile of 2 days. Applying this threshold, the weighted prevalence of prolonged LOS was 28.2%. Readmission was 5.1%. The first radiological assessment did not prove objective response (OR) in 31.5% of patients.

**Textbook outcome**

Components of TO and its final achievement are graphed in the *Figure 1*. The most frequent quality indicators met was the absence of mortality (97.6%) and the absence of readmission (94.9%). The achievement of the objective radiological response (68.5%) and the avoidance of any complications (50.0%) were the indicators most difficult to achieve. The weighted prevalence of no prolonged LOS was 71.8%. These indicators determined a final TO achievement in 29.0% of patients (95%C.I.: 20.4–39.5; I2: 90.9%).

The likelihood of achieving a TO (*Table 3*) was reduced in females (OR: 0.45; 95%C.I.: 0.30-0.67), when the ECOG PS was 1 (OR: 0.67; 95%C.I: 0.46-0.97), in presence of >3 tumors (OR: 0.46; 95%C.I: 0.29-0.70) and larger tumors (OR: 0.83; 95%C.I: 0.78-0.89). Detailed results for these four prognostic factors in determine each component of TO are in the *Supplementary Table 3*. Briefly, females had higher PES and prolonged LOS than males, patients with PS of 1 had higher morbidity, mortality and prolonged LOS, and patients with larger and multinodular tumors had higher PES, longer LOS and lower radiological objective response.

**Risk-adjustment**

 The removal of the predictable center’s effects resulting from the multilevel mixed-effect model provided the prediction of what each center's outcome would have been for the same standard population. The resulting risk-adjustment is reported in the *Figure 2*. The predicted TO achievement was 27.7% on the weighted average, and all centers involved had values adjacent to this estimation. This indicates that when adjusted for gender, ECOG and tumor burden, the outcome was similar across involved centers without any center’s volume effect.

 A tool for external comparison was built using the linear predictor of the fixed part of the model adding the random effect generated by each participating center. Mathematical details are provided in the *Supplementary Table 4*. Briefly, this tool is used by entering the number of patients in the external cohort, the number of patients who achieved TO, together with the percentages of females, of patients with ECOG-PS1, of patients with more than 3 tumors and the mean of the largest diameter. The tool will compare the observed events with the expected calculated by the model providing a risk adjusted TO achievement which can be compared with the present benchmark of 27.7%. A web calculator can be found at: <https://jscalc.io/calc/bz7p3W8LgH21CuDy>.

**Determinants of survival**

 During a median follow-up of 19 months, 639 patients died (56.9%). The median overall survival of the entire cohort was 26.2 months (95%C.I.: 24.6-28.7), with a 1-, 2- and 3-year survival rates of 78.5%, 55.0% and 36.0%, respectively. In the time-dependent analysis, patients achieving TO have a median overall of 35.4 months (95%C.I.: 31.8-41.9) and that of patients not achieving TO was 24.1 months (95%C.I.: 21.3-25.3). This resulted in an HR of 0.57 (95%C.I.: 0.44-0.72). When TO effect on survival was tested with the remaining baseline clinical features (*Table 4*) the HR remained 0.57 (95%C.I.: 0.46-0.70) and other favorable prognostic factors for survival were the HCV positivity (HR:0.60; 95%C.I.: 0.44-0.83) and the Child – Pugh class A (HR:0.60; 95%C.I.: 0.47-0.77).

**DISCUSSION**

Optimal outcome after TACE embraces an uneventful post-procedural course and the achievement of radiological response. With the present study, we provided such a composite endpoint in a large patient population from different centers for the first time to date. We observed that TO across centers was highly variable, being determined by the different clinical and radiological features of treated patients in each of them. After risk-adjustment we observed that this variability mainly depends on gender, ECOG PS and tumor burden and that, as predictable, patients achieving a TO had prolonged overall survival. All these aspects deserve appropriate discussion.

 The very first aspect highlighted in the present study is represented by the considerable heterogeneity of included patients. Despite restrictive inclusion criteria, which excluded patients with ECOG PS of 2, those with Child – Pugh score >7, each center treated very different patients, with the adoption of DEB-TACE in some of them and the assistance of CT scan during the procedure. All these aspects robustly support the usual observation that the stage in which TACE is applied encompass a great variety of patients, with different tumoral and clinical features *[Semin Liver Dis. 2012 Nov;32(4):348-59; Hepatology. 2020 Jul; 72(1): 198–212; Hepatology. 2015 Mar;61(3):905-14.]*. This a-priori concept was the reason why we adopted a weighted analysis rather than consider all patients as a unique study population, finally providing an estimation of such heterogeneity. In fact, to the best of our knowledge, an estimation of heterogeneity was never provided outside meta-analyses and, more importantly, was never provided for each clinical feature characterizing the patient undergoing TACE.

 The second important result regards details on adverse events occurred. Lots of data are available in literature, being this knowledge most from single center experiences *[Radiology. 2019;290(1):254-261; Curr Med Imaging Rev. 2019;15(4):380-385; HPB (Oxford). 2015;17(12):1137-1144]* or RCTs *[Br J Cancer. 2014;111(2):255-264; Cardiovasc Intervent Radiol. 2010;33(1):41-52]* with the impossibility to assess heterogeneity. Even meta-analyses available never produced the present in-deep estimation *[J Dig Dis. 2016;17(8):510-517; Dig Liver Dis. 2016;48(6):571-577]*. Here we observed that the weighted mortality was 2.4%. The null heterogeneity confirmed this as a robust outcome indicator. A dedicated discussion should be reserved to PES which was the most frequent adverse event, being of grade 1 or 2 in 24.7% of patients. The large heterogeneity observed (94.6%) is the consequence of at least three factors: 1) the subjective susceptibility to post-procedural pain of each patient, 2) the accuracy of clinicians in diagnose PES and 3) the different clinical characteristics of treated patients. When investigating this last aspect, we observed an unexpected correlation with the female gender. In the surgical setting, it is largely recognized that females have more frequently post-operative nausea and vomiting than males *[Anesth Essays Res. 2016 Sep-Dec; 10(3): 388–396; Anesthesiology. 2003 Jan;98(1):46-52]*. This is the consequence of both anesthetic drugs as well as surgical injury. TACE can resemble a surgical injury which in turn can determine higher susceptibility in females. Additionally, females may be more susceptible to sedation with fentanyl *[Anesth Essays Res. 2016 Sep-Dec; 10(3): 388–396; Anesthesiology. 2003 Jan;98(1):46-52]*, so that since nausea and vomiting are components of PES, the observed correlation found a reasonable justification. On this background, further dedicated studies, focused on components of PES in relationship with gender and sedation, are warranted.

 Other correlations between clinical features, TO and its components were somewhat expected. Patients with ECOG PS 1 were more likely to incur in post-TACE morbidity and mortality resulting in prolonged in-hospital stay and hospital readmission. These patients are fundamentally more fragile subjects which are more susceptible to TACE-induced liver injury. Morbidity was also determined by the number and the dimension of treated lesions. Being the largest area to treat, the more pronounced is the injury, as well as higher the dose of doxorubicin administered. Regarding this last aspect, we observed that DEB-TACE only apparently determined an increased likelihood to achieve the TO, whereas it was no more when adjusting for potential confounders. In particular, we did not observe DEB-TACE to be a prognostic factor for a reduction in PES (additional data: OR: 0.84; 95%C.I.:0.61 – 1.14) and the present observation confirms other important studies and RCTs which suggest the absence of a superiority of DEB-TACE over cTACE *[Radiology. 2019 Jan;290(1):254-261; Br J Cancer. 2014;111(2):255-264; Cardiovasc Intervent Radiol. 2010;33(1):41-52; Dig Liver Dis. 2016;48(6):571-57]*. Finally, it was also expected that OR was a function of diameter and number of treated nodules. Overall, all the present findings must be considered as clinically robust and suitable for TO assessment.

 All heterogeneities observed across centers were then reduced through risk-adjustment analysis. As noticeable, the TO prevalence after this adjustment was very similar among participating centers, being not related to the amount of cases provided to the present database, or to the volume of the radiology involved. This observation suggests that TACE is now a standardized procedure, which was effectively and properly applied in each center involved, which simply adapt their indications in relation to the volume of procedures they perform.

 Finally, TO achievement was an independent predictor of improved survival, thus satisfying the necessary early quality indicator of long-term prognosis. This was the consequence of predictors of TO achievement, which included both performance status and tumor characteristics as well as its definition, which included the achievement of an objective radiological response. In addition to TO, the Child – Pugh class and the HCV status determined the overall survival. While the impact on survival of more advanced liver dysfunction is intuitive, the favorable course of HCV positive patients is more intriguing but data supporting this finding are already provided. In patients treated with sorafenib, HCV positivity is a recognized predictor of better survival *[Future Oncol. 2019;15:3411-3422. doi: 10.2217/fon-2019-0287; J Clin Oncol. 2017;35(6):622-628. doi: 10.1200/JCO.2016.69.5197.]*. This can be explained from the molecular standpoint, where non-proliferation class HCC is more frequently observed in HCV patients *[Nat Rev Clin Oncol. 2018 Oct;15(10):599-616]*. However, even in patients with untreated HCC a remarkable survival difference was observed between western (North American and European) and eastern (Asia-Pacific) studies, with the first being affected by a higher prevalence of HCV positive patients and resulting in improved overall survival *[Hepatology. 2010 Apr;51(4):1274-83]*. In addition, the recent availability of an effective etiological treatment for HCV infection, represented by direct-acting antiviral agents (DAAs), could be a further reason for our findings. It has been showed that DAAs improve survival after locoregional treatments, including TACE *[Hepatology. 2020 Jun;71(6):1910-1922.]*, probably through a reduction of the risk of hepatic decompensation *[J Hepatol. 2019 Aug;71(2):265-273]*. Although data on HCV-RNA positivity and/or antiviral treatments were unfortunately unavailable, it is expected that a proportion of the patients in our cohort could have obtained DAA-induced sustained virologic response, as they were enrolled between 2014 and 2018. On the other hand, HCV-IgG negative patients are represented by patients with chronic hepatitis B, or patients with alcoholic or non-alcoholic fatty liver disease which already demonstrated have a worst survival in historical cohorts *[Hepatology. 2020 Jul;72(1):198-212.]*.

 Limitations of the present study are common to all the retrospective studies. The most affected aspects are probably represented by the assessment of PES, since the grade assessment can be biased if not defined prior to its registration. This probably represent a cause of the heterogeneity observed. However, as previously stated, PES is very subjective when pain or nausea are the only clinical symptoms reported, so that even under a prospective approach, the heterogeneity in reporting it should not change particularly. In addition, for the main outcome endpoint, the quality indicator of complication was defined as “any complication” so that PES reporting bias would be probably minimized. Of sure, a more structured approach to PES registration, through eventual appropriate tools such as Visual Analog Scale. Further appropriate studies investigating this aspect are warranted. Another issue is represented by the assessment of mRECIST, which is included in the TO definition. The obvious lack of central radiological review weakens the response evaluation. However, as suggested by the Food and Drugs Administration Clinical Trial Imaging Endpoint Process Standards 2018 *[ref]*, centralized image interpretation is not always critical when using some aspects of quantitative imaging when this measure is widely performed and reported in clinical medicine, as for mRECIST evaluation. The present approach certainly introduces more uncertainty *[J Hepatol. 2020 Aug;73(2):342-348.]* but represents a valuable snapshot of the common clinical practice.

 In conclusion, we here provided for the first time a risk-adjusted assessment of TO as quality indicator after TACE for HCC. The great heterogeneity among clinical features of enrolled patients produces different TO probabilities which mostly depend on gender, performance status and tumor burden. The TO achievement provides prolonged overall survival. Present results can be used for both healthcare quality assessment and for patient information regarding the likelihood of the success of the treatment.

**Table 1.** Clinical features of the study population of 1124 patients treated with TACE for HCC.

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Unweighted values** | **Weighted values (95%C.I.)** | **Heterogeneity (I2)** |
| Age (years) | 68.1 ± 9.9 | 68.9 (67.1 – 70.7) | 89.8% |
| Male | 905 (80.5%) | 80.0% (76.4 – 83.1) | 45.0% |
| HBsAg positive | 125 (11.1%) | 12.0% (9.1 – 15.8) | 61.7% |
| HCV- IgG positive | 361 (32.1%) | 30.5% (19.1 – 45.0) | 94.6% |
| Alcohol | 354 (31.5%) | 28.2% (20.0 – 38.2) | 89.9% |
| DEB-TACE | 502 (44.7%) | 41.2% (20.9 – 65.0) | 95.9% |
| Cone-beam TACE | 169 (15.0%) | 13.4% (5.0 – 31.2) | 88.9% |
| ECOG - PS 1 | 267 (23.8%) | 20.4% (12.8 – 31.0) | 91.3% |
| Child – Pugh B7 | 119 (10.6%) | 11.4% (8.2 – 15.7) | 69.2% |
| MELD score >10 | 372 (33.1%) | 33.0% (26.0 – 41.0) | 84.3% |
| ALBI score >1 | 695 (61.8%) | 66.2% (54.2 – 76.4) | 92.7% |
| Single lesion | 444 (39.5%) | 37.2% (30.2 – 44.9) | 83.3% |
| Two or three lesions | 463 (41.2%) | 42.8% (37.3 – 48.5) | 69.6% |
| More than three lesions | 217 (19.3%) | 17.5% (10.8 – 27.2) | 91.5% |
| Largest diameter (cm) | 4.1 ± 2.7 | 4.0 (2.4 – 5.7) | 95.0% |

Unweighted values derive from considering all patients from different centers as a unique study population. Weighted values and heterogeneity (I2) derived from meta-analysis of each single center values through the DerSimonian-Laird estimator. Detailed data stratified by participating center are provided in the *Supplementary Table 1*.

I2 statistic can be interpreted as follows: values of <25% =low heterogeneity; 25% - 50% = medium, 51% - 75% = substantial and >75% = considerable heterogeneity.

Abbreviations: HCV = hepatitis C virus, DEB = drug-eluting beads; CT = computed tomography, ECOG = Eastern Cooperative Oncology Group; PS = performance status, MELD = model for end-stage liver disease; ALBI = albumin-bilirubin.

**Table 2.** Complications after TACE in the study population of 1124 HCC patients.

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Unweighted values** | **Weighted values (95%C.I.)** | **Heterogeneity (I2)** |
| Post-procedural death | 21 (1.9%) | 2.4% (1.5 – 3.4) | 0.0% |
| Any complication | 594 (52.9%) | 50.0% (36.1 – 63.9) | 94.4% |
| Post-embolization syndrome |  |  |  |
| Grade 1-2 | 301 (26.8%) | 24.7% (14.4 – 39.0) | 94.6% |
| Grade 3-4 | 98 (8.7%) | 7.2% (3.9 – 12.9) | 83.7% |
| Liver dysfunction |  |  |  |
| Grade 1-2 | 237 (21.1%) | 13.2% (6.3 – 25.7) | 94.4% |
| Grade 3-4 | 48 (4.3%) | 2.6% (1.0 – 7.0) | 85.3% |
| Inguinal hematoma |  |  |  |
| Grade 1-2 | 22 (2.0%) | 2.5% (1.7 – 3.8) | 0.0% |
| Grade 3-4 | 1 (0.1%) | 0.7% (0.2 – 1.5) | 0.0% |
| Acute kidney injury |  |  |  |
| Grade 1-2 | 14 (1.3%) | 2.0% (1.3 – 3.5) | 9.3% |
| Grade 3-4 | 2 (0.2%) | 0.6% (0.3 – 1.3) | 0.0% |
| Acute cholecystitis |  |  |  |
| Grade 1-2 | 7 (0.6%) | 1.2% (0.6 – 2.2) | 0.0% |
| Grade 3-4 | 2 (0.2%) | 0.7% (0.3 – 1.5) | 0.0% |
| Acute pancreatitis |  |  |  |
| Grade 1-2 | 1 (0.1%) | 0.7% (0.3 – 1.6) | 0.0% |
| Grade 3-4 | 0 (0.0%) | 0.5% (0.2 – 1.3) | 0.0% |
| Liver abscess |  |  |  |
| Grade 1-2 | 2 (0.2%) | 0.8% (0.4 – 1.7) | 0.0% |
| Grade 3-4 | 5 (0.4%) | 0.9% (0.4 – 1.7) | 0.0% |
| Gastrointestinal bleeding |  |  |  |
| Grade 1-2 | 3 (0.3%) | 0.9% (0.5 – 1.9) | 0.0% |
| Grade 3-4 | 1 (0.1%) | 0.6% (0.3 – 1.3) | 0.0% |
| Prolonged in-hospital stay\*  | 228 (20.3%) | 28.2% (16.9 – 43.2) | 95.0% |
| Readmission within 30 days | 61 (5.4%) | 5.1% (3.0 – 8.4) | 68.6% |
| No radiological response | 427 (38.0%) | 31.5% (21.0 – 44.5) | 93.6% |

Unweighted values derive from considering all patients from different centers as a unique study population. Weighted values and heterogeneity (I2) derived from meta-analysis of each single center values through random-effect model. Detailed data stratified by participating center are provided in the *Supplementary Table 2*.

I2 statistic can be interpreted as follows: values of <25% =low heterogeneity; 25% - 50% = medium, 51% - 75% = substantial and >75% = considerable heterogeneity.

Continuity correction of 0.5 in studies with zero cell frequencies was adopted.

\* Defined as defined as a postoperative stay ≤75th percentile of the median values from the total cohort,

**Table 3.** Multilevel mixed-effect models on determinants of TO across participating centers.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Simple regression** |  | **Multivariable regression** |
| **Characteristic** | **OR (95%C.I.)** |  | **OR (95%C.I.)** |
| Age (years) | 0.99 (0.98 – 0.99) |  | - |
| Female | 0.43 (0.29 – 0.64) |  | 0.45 (0.30 – 0.67) |
| HBsAg positive | 0.92 (0.59 – 1.42) |  | - |
| HCV- IgG positive | 0.73 (0.52 – 1.01) |  | - |
| Alcohol | 1.57 (1.16 – 2.13) |  | 1.33 (0.88 – 2.01) |
| DEB-TACE | 1.53 (1.01 – 2.31) |  | 1.32 (0.88 – 1.98) |
| Cone-beam TACE | 1.13 (0.64 – 2.01) |  | - |
| ECOG - PS 1 | 0.60 (0.42 – 0.87) |  | 0.67 (0.46 – 0.97) |
| Child – Pugh B7 | 1.02 (0.65 – 1.59) |  | - |
| MELD score >10 | 0.90 (0.66 – 1.21) |  | - |
| ALBI score >1 | 0.76 (0.57 – 1.02) |  | - |
| Single lesion | Ref. |  | Ref. |
| Two or three lesions | 1.05 (0.78 – 1.42) |  | Ref. |
| More than three lesions | 0.46 (0.29 – 0.70) |  | 0.46 (0.31 – 0.69) |
| Largest diameter (cm) | 0.81 (0.76 – 0.86) |  | 0.83 (0.78 – 0.89) |

The residual intraclass correlation after multivariable analysis was 0.139 (95%C.I.: 0.56 – 0.31) meaning that conditional on the fixed-effects covariates, the center effect accounts for approximately 13.9% of the total residual variance.

The final mixed-effect logistic regression had an AUC of 0.751 (95%C.I.: 0.72 – 0.78) and the corresponding plot is reported in the *Supplementary Figure 1*.

**Table 4.** Results from Cox regression with TO as time-dependent covariate, on overall survival.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Simple regression** |  | **Multivariable regression\*** |
| **Characteristic** | **HR (95%C.I.)** |  | **HR (95%C.I.)** |
| TO achievement | 0.57 (0.44 – 0.72) |  | 0.57 (0.46 – 0.70) |
| Age (years) | 1.01 (0.99 – 1.02) |  | - |
| Female \* | 0.96 (0.74 – 1.23) |  | Not evaluated |
| HBsAg positive | 0.88 (0.61 – 1.27) |  | - |
| HCV- IgG positive | 0.61 (0.46 – 0.80) |  | 0.60 (0.44 – 0.83) |
| Alcohol | 1.23 (0.98 – 1.51) |  | - |
| DEB-TACE | 1.32 (0.76 – 2.28) |  | - |
| Cone-beam TACE | 0.74 (0.44 – 1.21) |  | - |
| ECOG - PS 1 \* | 1.41 (1.06 – 1.92) |  | - |
| Child – Pugh B7 † | 1.58 (1.23 – 2.03) |  | 1.66 (1.29 – 2.11) |
| MELD score >10 | 1.34 (0.94 – 1.86) |  | - |
| ALBI score >1 † | 1.37 (1.07 – 1.75) |  | Not evaluated |
| Single lesion | Ref. |  | Not evaluated |
| Two or three lesions | 1.31 (1.09 – 1.58) |  | Not evaluated |
| More than three lesions \* | 1.84 (1.19 – 2.85) |  | Not evaluated |
| Largest diameter (cm) \* | 1.14 (1.09 – 1.20) |  | Not evaluated |

\* These factors were related to Textbook Outcome (*Table 3*) and not entered the multivariable analysis because of the observed collinearity. The HRs resulting from simple regression were provided for descriptive purposes only.

† Child – Pugh class B was chosen instead of ALBI score because of the lower -2 Log-likelihood.

**LEGENDS TO FIGURES**

**Figure 1.** Proportion of patients estimated through random-effect modelling for whom each desired health outcome forming the Textbook Outcome (TO) was realized.

**Figure 2.** TO estimation after risk-adjustment for each center involved. As noticeable, all the centers had TO values around the average of 25.5% (dotted line). Darker columns represent centers with a volume of >200 procedures/years, brighter columns represent centers with a volume of <100 procedures/years. Intermediate gray columns represent centers with a volume of 100-200 procedures/years. Bars represent 95% confidence intervals

**FIGURE 1.**

****

**FIGURE 2.**

 ****

**APPENDIX**

Adverse events definitions:

|  |  |  |
| --- | --- | --- |
| **AE Grading** | **Grades 1-2** | **Grades 3-4** |
| **Post Embolization Syndrome (PES)** | Mild pain, with or without fever, with or without nausea / vomiting requiring eventual administration of analgesic, antiemetics, antipyretics. | Moderate to severe pain, with mild to moderate peritonitis requiring narcotics, fever requiring antibiotics, vomiting requiring naso-gastric tube placement |
| **Cholecystitis (Acute)** | Asymptomatic with radiographic findings only/Symptomatic requiring medical intervention | Interventional radiology, endoscopic, or operative intervention indicated / Life-threatening consequences (i.e. sepsis or perforation) |
| **Pancreatitis (Acute)** | Asymptomatic, enzyme elevation and/or radiographic findings or symptomatic, medical intervention indicated | Interventional radiology or operative intervention indicated / Life-threatening consequences (i.e. circulatory failure, hemorrhage, sepsis) |
| **Liver Abscess** | Radiological finding only without any clinical signs of systemic infection. Analgesics and electrolytes allowed  | Antibiotic, antifungal intervention indicated; radiologic or surgical intervention indicated / Life-threatening consequences  |
| **Acute kidney injury** | Creatinine 2-3 x above baseline value | Creatinine >3 x baseline or >4.0mg/dL or requiring dialysis |
| **Inguinal hematoma** | Mild symptoms / minimally invasive evacuation or aspiration indicated, medical therapy | Transfusion, radiologic, endoscopic or intervention indicated |
| **Liver dysfunction** | Transient bilirubin increases ≥ 50% compared with baseline, Albumin decrease by 0.3 g/dL, AST or ALT increase of > 25%, INR increase of ≥ 25%, eventual albumin supplementation for ascites | Development of jaundice, encephalopathy, increasing ascites eventually requiring drainage / Life-threatening consequences |
| **GI bleeding** | Hematemesis or tarry stool without the need for transfusion or endoscopy | Transfusion, radiologic, endoscopic or intervention indicated / Life-threatening consequences |

**Supplementary Table 1.** Main clinical and radiological features by each center included in the analysis.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Birmingham (n=261) | Mainz (n=166) | Freiburg (n=164) | Bologna (n=154) | Palermo (n=97) | Nottingham (n=90) | Houston (n=72) | Milan(n=55) | Klagenfurt (n=35) | Faenza (n=30) |
| Age (years) | 66.2 ± 9.6 | 66.9 ± 9.3 | 68.4 ± 10.4 | 67.9 ± 10.6 | 72.4 ± 8.6 | 71.1 ± 7.5 | 68.9 ± 9.2 | 63.1 ± 11.1 | 73.7 ± 6.3 | 70.0 ± 12.3 |
| Male | 218 (83.5%) | 142 (85.5%) | 138 (84.1%) | 114 (74.0%) | 69 (71.1%) | 72 (80.0%) | 56 (77.8%) | 46 (83.6%) | 26 (74.3%) | 24 (80.0%) |
| HBsAg positive | 15 (5.7%) | 15 (9.0%) | 22 (13.4%) | 24 (15.6%) | 7 (7.2%) | 10 (11.1%) | 11 (15.3%) | 10 (18.2%) | 3 (8.6%) | 8 (26.7%) |
| HCV-IgG positive | 60 (23.0%) | 30 (18.1%) | 41 (25.0%) | 85 (55.2%) | 77 (79.4%) | 8 (8.9%) | 20 (27.8%) | 28 (50.9%) | 4 (11.4%) | 8 (26.7%) |
| Alcohol | 98 (37.5%) | 85 (51.2%) | 45 (27.4%) | 36 (23.4%) | 5 (5.2%) | 35 (38.9%) | 6 (8.3%) | 16 (29.1%) | 22 (62.9%) | 6 (20.0%) |
| DEB-TACE | 71 (27.2%) | 131 (78.9%) | 111 (67.7%) | 2 (1.3%) | 0 (0.0%) | 55 (61.1%) | 71 (98.6%) | 55 (100%) | 0 (0.0%) | 6 (20.0%) |
| CT- TACE | 0 (0.0%) | 34 (20.5%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 8 (8.9%) | 72 (100%) | 13 (23.6%) | 35 (100%) | 7 (23.3%) |
| ECOG - PS 1 | 107 (41.0%) | 1 (1.8%) | 38 (23.2%) | 13 (8.4%) | 35 (36.1%) | 24 (26.7%) | 30 (41.7%) | 1 (1.8%) | 0 (0.0%) | 7 (23.3%) |
| Child – Pugh B7 | 11 (4.2%) | 25 (15.1%) | 16 (9.8%) | 18 (11.7%) | 17 (17.5%) | 13 (14.4%) | 2 (2.8%) | 4 (7.3%) | 8 (22.9%) | 5 (16.7%) |
| MELD score >10 | 73 (28.0%) | 66 (39.8%) | 55 (33.5%) | 76 (49.4%) | 26 (26.8%) | 11 (12.2%) | 13 (18.1%) | 19 (34.5%) | 16 (45.7%) | 17 (56.7%) |
| ALBI score >1 | 113 (43.3%) | 136 (81.9%) | 78 (47.6%) | 117 (76.0%) | 79 (81.4%) | 60 (66.7%) | 33 (45.8%) | 27 (49.1%) | 32 (91.4%) | 20 (66.7%) |
| Single lesion | 141 (54.0%) | 39 (23.5%) | 59 (36.0%) | 71 (46.1%) | 27 (27.8%) | 42 (46.7%) | 23 (31.9%) | 18 (32.7%) | 11 (31.4%) | 13 (43.3%) |
| 2-3 lesions | 103 (39.5%) | 48 (28.9%) | 64 (39.0%) | 61 (39.6%) | 58 (59.8%) | 37 (41.1%) | 35 (48.6%) | 28 (50.9%) | 18 (51.4%) | 11 (36.7%) |
| >3 lesions | 17 (6.5%) | 79 (47.6%) | 41 (25.0%) | 22 (14.3%) | 12 (12.4%) | 11 (12.2%) | 14 (19.4%) | 9 (16.4%) | 6 (17.1%) | 6 (20.0%) |
| Largest diameter (cm) | 4.7 ± 2.8 | 3.9 ± 2.7 | 4.8 ± 3.0 | 2.7 ± 1.3 | 3.1 ± 1.4 | 5.4 ± 3.4 | 4.1 ± 2.8 | 3.2 ± 1.4 | 4.4 ± 1.9 | 4.0 ± 3.4 |
| #TACE /year | 100-200 | >200 | >200 | >200 | >200 | 100-200 | 100-200 | 100-200 | <100 | <100 |

**Supplementary Table 2.** Clinical outcomes by each center included in the analysis.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Birmingham (n=261) | Mainz (n=166) | Freiburg (n=164) | Bologna (n=154) | Palermo (n=97) | Nottingham (n=90) | Houston (n=72) | Milan(n=55) | Klagenfurt (n=35) | Faenza (n=30) |
| Post-procedural death | 4 (1.5%) | 1 (0.6%) | 5 (3.0%) | 2 (1.3%) | 2 (2.1%) | 4 (4.4%) | 0 (0.0%) | 0 (0.0%) | 2 (5.7%) | 1 (3.3%) |
| Any complication | 167 (64.0%) | 95 (57.2%) | 137 (83.5%) | 17 (11.0%) | 45 (46.4%) | 39 (43.3%) | 27 (37.5%) | 38 (69.1%) | 11 (31.4%) | 18 (60.0%) |
| PES G1-G2 | 38 (14.6%) | 54 (32.5%) | 98 (59.8%) | 10 (6.5%) | 17 (17.5%) | 14 (15.6%) | 26 (36.1%) | 35 (63.6%) | 1 (2.9%) | 10 (33.3%) |
| G3 – G4 | 14 (5.4%) | 29 (17.5%) | 33 (20.1%) | 3 (1.9%) | 0 (0.0%) | 9 (10.0%) | 2 (2.8%) | 0 (0.0%) | 8 (22.9%) | 0 (0.0%) |
| Dysfunction G1-G2 | 135 (51.7%) | 4 (2.4%) | 11 (6.7%) | 2 (1.3%) | 34 (35.1%) | 22 (24.4%) | 0 (0.0%) | 15 (27.3%) | 2 (5.7%) | 12 (40.0%) |
| G3 – G4 | 5 (1.9%) | 4 (2.4%) | 32 (19.5%) | 0 (0.0%) | 0 (0.0%) | 4 (4.4%) | 0 (0.0%) | 0 (0.0%) | 2 (5.7%) | 1 (3.3%) |
| Hematoma G1-G2 | 6 (2.3%) | 8 (4.9%) | 2 (1.2%) | 0 (0.0%) | 1 (1.0%) | 2 (2.2%) | 1 (1.4%) | 0 (0.0%) | 1 (2.9%) | 1 (3.3%) |
| G3 – G4 | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (2.9%) | 0 (0.0%) |
| Kidney injury G1-G2 | 0 (0.0%) | 3 (1.8%) | 5 (3.0%) | 0 (0.0%) | 2 (2.1%) | 4 (4.4%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| G3 – G4 | 1 (0.4%) | 1 (0.6%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| Cholecystitis G1-G2 | 0 (0.0%) | 0 (0.0%) | 2 (1.2%) | 3 (1.9%) | 1 (1.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (3.3%) |
| G3 – G4 | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (1.1%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| Pancreatitis G1-G2 | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 1 (3.3%) |
| Liver abscess G1-G2 | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 2 (2.1%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| G3 – G4 | 0 (0.0%) | 1 (0.6%) | 1 (0.6%) | 0 (0.0%) | 0 (0.0%) | 2 (2.2%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| GI bleeding G1-G2 | 0 (0.0%) | 0 (0.0%) | 3 (1.8%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| G3 – G4 | 0 (0.0%) | 0 (0.0%) | 1 (0.6%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| Prolonged LOS  | 36 (13.8%) | 13 (37.1%) | 56 (34.1%) | 61 (39.6%) | 24 (24.7%) | 9 (10.0%) | 2 (2.8%) | 1 (1.8%) | 13 (37.1%) | 13 (43.3%) |
| Readmission | 13 (5.0%) | 4 (2.4%) | 21 (12.8%) | 3 (1.9%) | 4 (4.1%) | 10 (11.1%) | 2 (2.8%) | 0 (0.0%) | 1 (2.9%) | 3 (10.0%) |
| No OR  | 92 (35.2%) | 82 (49.4%) | 108 (65.9%) | 13 (8.4%) | 50 (51.5%) | 47 (52.2%) | 15 (20.8%) | 8 (14.5%) | 3 (8.6%) | 9 (30.0%) |
| TO achievement | 58 (22.2%) | 48 (28.9%) | 14 (8.5%) | 81 (52.6%) | 19 (19.6%) | 27 (30.0%) | 37 (51.4%) | 13 (23.6%) | 17 (48.6%) | 7 (23.3%) |

**Supplementary Table 3.** Results of mixed-effect logistic regression of gender, ECOG, number and diameter of tumors in determine TO components.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | PES (any grade) | Any adverse event | Mortality | Prolonged LOS | Readmission <30 days | Radiological Objective response |
|  | OR (95%C.I.) | OR (95%C.I.) | OR (95%C.I.) | OR (95%C.I.) | OR (95%C.I.) | OR (95%C.I.) |
| Female | 1.67 (1.17 – 2.38) | 1.94 (1.35 – 2.77) | 0.95 (0.31 – 2.89) | 1.80 (1.21 – 2.59) | 0.92 (0.46 – 1.83) | 0.84 (0.89 – 1.18) |
| ECOG PS1 | 1.36 (0.95 – 1.97) | 1.51 (1.08 – 2.11) | 3.41 (1.37 – 8.47) | 1.73 (1.17 – 2.54) | 1.72 (0.97 – 3.04) | 0.83 (0.60 – 1.15) |
| More than 3 lesions | 1.86 (1.28 – 2.69) | 2.36 (1.61 – 3.43) | 0.77 (0.22 – 1.22) | 0.61 (0.38 – 0.96) | 1.28 (0.66 – 2.48) | 0.39 (0.27 – 0.55) |
| Largest tumor diameter (per cm) | 1.19 (1.12 – 1.26) | 1.20 (1.13 – 1.28) | 1.08 (0.96 – 1.21) | 1.09 (1.03 – 1.16) | 0.99 (0.90 – 1.09) | 0.89 (0.84 – 0.94) |

**Supplementary Table 4.** Instruction to compute the comparison between an external center and the present risk-adjusted analysis.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **STEP 1. Calculation of Expected Probabilities**1. Compute the linear predictor of the fixed part of the model:

xb = 1.916 if male + 0.634 if ECOG-PS of 1 + 0.441 if >3 lesions + 0.800 per cm of largest diameter;In this formula, data are entered as proportions (male, ECOG-PS1 and lesions>3) and mean (tumor diameter) for the external cohort.1. Add the following values to solve the subsequent formula:

|  |  |  |
| --- | --- | --- |
| Center | Random effect values | n |
| #1 | -0,81488 | 55 |
| #2 | -1,61516 | 164 |
| #3 | 0,360922 | 154 |
| #4 | -0,93586 | 97 |
| #5 | 0,473366 | 35 |
| #6 | -0,30496 | 166 |
| #7 | -0,63594 | 261 |
| #8 | 0,725746 | 72 |
| #9 | -0,0848 | 90 |
| #10 | -0,57069 | 30 |

1. The weighted linear predictor is then calculated as follows: Wxb = [(xb + re1) \* n1 + (xb + re2) \* n2 + …. (xb + re10) \* n10] / 1124;
2. Calculate the expected weighted probability of TO across participating centers as follows: Expected = 1 / (1 + e^-Wxb)

**STEP 2. Calculation of observed-to-expected ratio**1. Calculate the observed TO proportion in your external cohort as Observed = TOexternal / nexternal
2. Calculate the observed-to-expected ratio as SToR = Observed / Expected
3. The risk adjusted TO value (RSToR) is finally calculated by the following formula: (321 + TOexternal / 1124 + nexternal) \* SToR.
4. Compute the 95%C.I. of SToR using nexternal.
5. If the 95%C.I. includes the present risk-adjusted value of 27.7%, then the external cohort had similar performance of the present cohort when covariate distribution was similar. If the 95%C.I. places over 27.7% it means that the external cohort performs better than the present cohort. If the 95%C.I. places below 27.7% it means that the external cohort performs worse than the present cohort.
 |

**Supplementary Figure 1.** Calibration plot for the final mixed-effect multivariable analysis model used for risk-adjustment across participating centers.

