Editorial

Resection or ablation for very early hepatocellular carcinoma and the fundamental problem of causal inference

Alessandro Cucchetti, Matteo Serenari

Department of Medical and Surgical Sciences, DIMEC, S.Orsola-Malpighi Hospital, Alma Mater Studiorum, University of Bologna, Bologna, Italy

Correspondence to: Alessandro Cucchetti, MD. Department of Medical and Surgical Sciences, DIMEC, Alma Mater Studiorum, University of Bologna, Via Albertoni, 15, 40138 Bologna, Italy. Email: aleqko@libero.it.

Provenance: This is an invited Editorial commissioned by Editor-in-Chief Yilei Mao (Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China).

Comment on: Liu PH, Hsu CY, Hsia CY, *et al.* Surgical Resection Versus Radiofrequency Ablation for Single Hepatocellular Carcinoma ≤ 2 cm in a Propensity Score Model. Ann Surg 2016;263:538-45.

Submitted Mar 22, 2017. Accepted for publication Apr 19, 2017.

doi: 10.21037/hbsn.2017.05.03

View this article at: http://dx.doi.org/10.21037/hbsn.2017.05.03

Despite strict size definition (single ≤2 cm), very early stage hepatocellular carcinoma (HCC) still represents an heterogeneous stage, regarding, in particular, liver function (i.e., Child-Pugh and MELD scores or bilirubin levels and presence or absence of portal hypertension), tumor location (deep in the liver, adjacent to major biliary or vascular structures or located on the surface of the liver) and patient characteristics (i.e., age, absence or presence of comorbidities as well as patient treatment preference). Recommendations from EASL and AASLD guidelines suggest hepatic resection (HR) when portal pressure and bilirubin are normal, that are, ideal surgical candidates, and ablation in presence of increased portal pressure/bilirubin and/or associated disease (1).

Existing literature cannot be considered of high-level when comparing HR to radiofrequency ablation (RFA) mainly due to study design, small sample sizes and lack of solid subgroup analyses (1,2). Only retrospective observational studies are available, aimed at identifying which treatment better works in a real clinical scenario. Propensity score (PS) match, like that published by Liu and colleagues in the issue of March 2016 of *Annals of Surgery* (3), represents a statistical technique that attempts to estimate the effect of a treatment, by accounting for the covariates that predict receiving the treatment, in the attempt to compare more homogeneous groups of patients, treated either with HR or RFA. In other words, PS match allows to identify those patients having similar clinical and tumoral characteristics but that were treated with HR or RFA, excluding the "outliers" from the analysis. This approach is effectively very similar to the procedure of randomized controlled trial (RCT) design when establishing inclusion criteria that should include baseline features common to patients eligible to both HR and RFA (4,5).

In this study, authors conclude that surgical resection provides better long-term overall survival (OS) and recurrence-free survival compared with RFA in patients with BCLC very early-stage HCC, and that surgical resection should be considered as the first-line treatment for these patients (2). While the superiority of HR over RFA in terms of RFS is well established, the finding of a survival benefit in terms of survival is more uncertain and one principal methodological concern, present in the Liu's study, supports this uncertainty. As expectable, in the un-matched cohort, patients submitted to RFA (n=128) were significantly older, with lower albumin levels and platelet count, higher ALT and INR and with mild lower tumor burden in respect to patients submitted to HR (n=109). The estimated 1-, 3-, and 5-year OS rates were 98%, 97%, and 81% after surgery and 98%, 88%, and 76% after RFA (P=0.136). After PS match, the estimated 1-, 3-, and 5-year OS rates were 97%, 97%, and 80% after surgery (n=79), thus, remained grossly unchanged, but surprisingly the corresponding figures after RFA (n=79) decreased to 97%, 83%, and 66%, respectively (P=0.034). This finding is rather counterintuitive, since the PS match should have excluded from the sub-analysis those older RFA patients with lower albumin levels and platelet count, higher ALT and INR, which represent

the "outliers" previously described, thus, those patients expected to go worse. The PS match would have probably led to a decreased sample size that instead was essential to create reliable comparable groups of patients, which, in turn, lead to confusion around prognostic factors. Findings from Liu and colleagues can resemble that reported from Lee and colleagues (6) which recently reported that RFA had better results than HR for Child-Pugh A5 patients with HCC within Milan criteria, while HR was superior for Child-Pugh > A5 patients, which was a counterintuitive finding that lacked face validity. Thus, it is likely that the analysis from Liu and colleagues suffers from some methodological bias.

It is clear that the estimation of the "causal effect" of some exposure regarding some outcomes must be the goal of all clinical studies; however, to measure the causal effect of HR over RFA, the outcome of both alternatives in the same individual would have to be known (7,8). However, it is impossible to observe both outcomes because one is unavoidably always lacking (counterfactual outcome). Even RCTs would not solve the problem of missing counterfactual cases since, in the attempt to minimize the impact of confounding factors, only patients fulfilling predefined characteristics are randomized. This selection excludes a considerable proportion of patients who would undergo the treatment in clinical practice. In addition, it is unlikely that the strict selection criteria used for RCTs could be adopted in the daily clinical practice.

It is unlikely that PS match or even RCT would fully depict the benefit, if any, that could be obtained from resect patients with very early HCC instead of ablation, but as the statistician Paul Holland said, "*science can make itself most valuable by measuring the effects of causes*" (8). The effect of a cause is the difference between what happens if some patients are exposed to some treatment against what would have been the result if they had not been (counterfactual outcomes). Whenever this aspect will be solved through dedicated statistical analyses (4,8), it will be still scarce the evidence supporting resection or ablation as best therapy to adopt for very early HCCs and the analysis from Liu and colleagues remains not sufficient to provide suggestions to change the current clinical guidelines for this tumor stage.

Cite this article as: Cucchetti A, Serenari M. Resection or ablation for very early hepatocellular carcinoma and the fundamental problem of causal inference. HepatoBiliary Surg Nutr 2017;6(4):272-273. doi: 10.21037/hbsn.2017.05.03

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Vitale A, Peck-Radosavljevic M, Giannini EG, et al. Personalized treatment of patients with very early hepatocellular carcinoma. J Hepatol 2017;66:412-23.
2. Cucchetti A, Piscaglia F, Cescon M, et al. Systematic review of surgical resection vs radiofrequency ablation for hepatocellular carcinoma. World J Gastroenterol 2013;19:4106-18.
3. Liu PH, Hsu CY, Hsia CY, et al. Surgical Resection Versus Radiofrequency Ablation for Single Hepatocellular Carcinoma ≤ 2 cm in a Propensity Score Model. Ann Surg 2016;263:538-45.
4. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res 2011;46:399-424.
5. Dahabreh IJ, Sheldrick RC, Paulus JK, et al. Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. Eur Heart J 2012;33:1893-901.
6. Lee YH, Hsu CY, Chu CW, et al. Radiofrequency ablation is better than surgical resection in patients with hepatocellular carcinoma within the Milan criteria and preserved liver function: a retrospective study using propensity score analyses. J Clin Gastroenterol 2015;49:242-9.
7. Hernán MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health 2004;58:265-71.
8. Holland PW. Statistics and causal inference (with discussion). J Am Stat Assoc 1986;81:945-61.

