**Appraisal of three studies (Lobo et al. (2020), Huang et al. (2010), Sharib et al. (2020)**

Three studies (Lobo et al. (2020), Huang et al. (2010), Sharib et al. (2020)) included in this review conducted cost-effectiveness analyses of management strategies for pancreatic cysts. Lobo et al. (2020) compared the 2017 International Consensus Guidelines (2017 ICG) with the 2015 American Gastroenterological Association guidelines (2015 AGA), Huang et al. (2010) compared the 2006 International Association of Pancreatology (IAP) guidelines with immediate surgery and no surveillance, and Sharib et al. (2020) evaluated the cost-effectiveness of the 2017 ICG with immediate surgery and no surveillance strategy.

All three studies aimed to evaluate the cost-effectiveness of different management strategies to determine whether higher sensitivity or specificity should be preferred. The three studies all adopted Markov model to conduct the analysis, but there exists heterogeneity of model parameters which resulted in varying model outcomes.

Regarding cost parameters, Lobo et al. (2020) considered only three components, namely imaging (MRI) cost, Whipple procedure cost, and distal pancreatectomy cost, which is different from two other studies. Huang et al. (2010) and Sharib et al. (2020), on the other hand, considered not only the costs directly associated with surgery and surveillance, but also the costs that may arise after surgery due to complications, until death. For example, Huang et al. (2010) and Sharib et al. (2020) considered the cost of physician visits, the cost of annual diabetes care, the cost of end-of-life care, the cost of metastatic treatment of cancer, and the cost of complications. These different cost components suggest that the study perspective may differ across studies, but this is not clearly stated. Only Huang et al. (2010) clearly stated that the study was based on the societal perspective, while the other two studies did not clearly state which perspective they chose.

Considering that Lobo et al. (2020) only considered the costs directly associated with surgery and surveillance, the costs might be underestimated because the long-term effects of management strategies were ignored. Huang et al. (2010) and Sharib et al. (2020) both estimated the long-term impact of treatment strategies, but costs may still be underestimated in these two studies as well because they did not include indirect costs to patients, i.e., loss of productivity due to illness. MRI costs are also heterogeneous across the three studies, ranging from $3,471 ($2,830-$4,431) to $544 ($250-$1,500). These heterogeneous cost parameters are one of the reasons contributing to the different model results.

Utilities applied to each health state of the model depends on the model complexity of each study, which also contributed to the varying model outcomes.

Huang et al. (2010) adopted simplified Markov model which did not stratify the dysplasia level. In this study, only three health states were considered -i.e., IPMN, invasive cancer, and death- and the differences in long-term health-related quality of life before and after surgery were considered almost negligible, so utility after the Whipple procedure was assumed to be 100%. However, Sharib et al (2020) stratified dysplasia into low and high levels and assumed that health utilities decreased during surgery (surgical disutility: -2.5%). Because both studies measured the long-term effect of management strategies, health utilities for surgical complications, diabetes, and metastatic cancer were all considered in the two studies. Lobo et al. (2020), in contrast, did not consider the utilities after the surgical procedure such as diabetes or metastatic cancer.

The three studies are similar in terms of objectives, but the cost-effectiveness results are different. This can be explained by the different perspective of the study and the heterogeneity of the model parameters. If the study estimates only the costs directly incurred by the management strategy such as Lobo et al. 2020, the costs are largely underestimated. The study results also differ by the model structure and assumptions underlying the utilities of health states. It is difficult to determine which factor is the driving force behind the different model results across the studies, but it is clear that heterogeneity in cost parameters contributed to the different cost results.