Point-by-Point Response

Stakeholder Comments for Health Technology Assessment

Clinical Effectiveness, Safety, Health Economic Characteristics and Benefit-Harm Balance of Treatments for Newly Diagnosed Metastatic Hormone-Sensitive Prostate Cancer

commissioned by the Swiss Medical Board

Assessment Team

Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich

Institute of Pharmaceutical Medicine (ECPM), University of Basel

07 July 2021

Stakeholder Comments

Full name	**Beat Kipfer**
Function	Vertrauensarzt kpt /Mitglied curafutura
Organization	**curafutura**

Comment number	Chapter	Comment	Suggested change	Response
General Comments				
1	Acknowledgements	• Die beigezogenen Experten sind für die Fachkompetenz in der Schweiz repräsentativ • Sowohl das CAT wie das HEAT sind erfahren und kompetent Die anderen Stakeholder waren ebenfalls ausgewogen vertreten (Industrie/Versicherung/Non-profit Sector)		Vielen Dank für Ihr detailliertes und weitgehend positives Feedback.
2	Executive Summary	PICO • Das PICO ist korrekt • Die Outcomes- Parameter sind repräsentativ für das prostata- Ca Man hätte als Kriterium noch den Gleason-Score beiziehen können für die Charakterisierung der Patientenkohorte (zusätzlich zu CHAARTED)		Der Gleason-Score ist ein wichtiges prognostisches Kriterium bei Prostata-Krebs. Der ursprüngliche Projektplan enthielt auch Subgruppen-Analysen für Low Risk/High Risk (nach LATITUDE-Klassifikation, welche den Gleason-Score mit einbezieht). Wie im Dokument beschrieben, haben wir letztlich aufgrund unzureichender Datenlage und niedriger zusätzlicher Aussagekraft bezüglich der

				Wirksamkeit der Behandlungen verzichtet. In Table 1 (Übersicht der Studien) zeigen wir den Anteil an Patienten mit hohem Gleason-Score (8-10) in den verschiedenen Studien auf.
3		Benefit-Harm Assessment • Die Parameter sind korrekt «Time horizon» für die Betrachtung ist korrekt		Keine Antwort notwendig.
4		Health Economic Literature Review • CHEERS ist adäquat Die Qualität der Studien im Hinblick auf die CHEERS – Kriterien war hingegen sehr heterogen (siehe S.100)		Keine Antwort notwendig.
5		Cost-Effectivness Analysis • Markov-Modell ist adäquat • Zeithorizont ist ebenfalls adäquat Discount-rate entspricht mit 3% dem Standard		Vielen Dank. Keine Antwort notwendig.
6		Budget Impact Analysis • Unklar, wie gross die Variabilität ist in den Schätzungen zwischen «neuen» mHSPC und Vorbehandelten, da dies aus nationalen Registern abgeleitet wurde.		Die Schätzung der Prävalenz und Inzidenz von mHSPC in der Schweiz, sowie der Marktanteile der verschiedenen Behandlungen in der Schweiz beinhalten gewisse Unsicherheiten. Aufgrund der Datenlage in der Schweiz waren leider nur grobe Schätzungen unter Zuhilfenahme von Daten aus

		• Unklar, wie die Verteilung auf die einzelnen Behandlungsarme aussieht («Marktanteile» von Abiraterone, Enzalutamide, Apalutamide und Docetaxel)		anderen Ländern möglich. Durch Szenario-Analysen haben wir diese Unsicherheiten exploriert. Eine Verbesserung der Datenlage in der Schweiz wäre aus unserer Sicht erstrebenswert.
7		Results and Discussion • Kleine Datenbasis (8 Trials, welche dir Kriterien erfüllten) • Heterogene Population in den Studien, divergierende alternative Therapien • HR für ADT + systemic mHSPC treatments zwischen 0.77 und 0.63, entsprechend klinisch relevant • Aussagen über health related quality of life ist sehr beschränkt möglich anhand der Studien • Gesundheitsökonomische Daten sind sehr heterogen geografisch. Die ICER sind wegen ihrer Unterschiedlichkeit zwischen den Studien schwierig zu interpretieren. Es zeigt sich jedoch ein Trend (Abiraterone mit dem höchsten ICER)		Vielen Dank für diese Evaluation.
Specific Comments				

8		Ein Auszug aus der SL: Apalutamide (Erleada): Aufnahme 01.11.2020, befristet bis 31.10.2022 • ERLEADA wird vergütet in Kombination mit Androgendeprivationstherapie (ADT) für die Behandlung von erwachsenen Patienten mit metastasiertem, hormonsensitivem Prostatakarzinom (mHSPC) Enzalutamide (Xtandi), Befristet bis 31.03.2023 • XTANDI wird vergütet in Kombination mit LHRH Agonisten zur Behandlung von Männern mit metastasiertem hormonsensitivem Prostatakarzinom (mHSPC) Abiraterone (Zytiga), keine Befristung • Zur Behandlung in Kombination mit Prednison oder Prednisolon (5mg/Tag) und Androgendeprivationstherapie (ADT) bei Patienten die innerhalb der letzten 3 Monate mit einem Hochrisiko metastasiertem hormonsensitivem Prostatakarzinom (mHSPC) neu diagnostiziert wurden.		Danke für diese Ausführung. Wir haben in der Overall Conclusion minimale Ergänzungen angebracht.

		• Als Hochrisiko wird das Vorliegen von mindestens 2 der 3 folgenden Risikofaktoren definiert: (1) Gleason-Score von ≥8; (2) Vorhandensein von mindestens 3 Läsionen im Knochenscan; (3) Vorhandensein messbarer viszeraler Metastasen (ohne Berücksichtigung des Lymphknotenbefalls).		
9		• Klare und eindeutige Darstellung der Effizienz der untersuchten Modalitäten, insbesondere auch Aufzeigen der Vor- und Nachteile der einzelnen Behandlungsarten • Einzelne Studien (ENZAMET) konnten relevant das Ergebnis beeinflussen (benefit.-harm balance für ADT+enzalutamide), hingegen beeinflusste dies nicht die Abgrenzung gegenüber der Kombination ADT+ docetaxel, welche das schlechteste BHA hatte. Zusammenfassend sind aber die Resultate angesichts der fehlenden Vergleichbarkeit mit Vorsicht zu interpretieren bezüglich BHA.		Vielen Dank für diese Evaluation.

Stakeholder Comments

Full name	**Marianne Eggenberger, Markus Gnaegi**
Function	
Organization	**santésuisse**

Comment number	Chapter	Comment	Suggested change	Response
General Comments				
1	1	The summary is very valuable and addresses the relevant issues. It gives a very good overview of the results of the study.	-	Many thanks for your valuable feedback.
2	5	The results found in the eight eligible RCTs are presented in a comprehensive and detailed manner in accordance with the predefined questions. At the same time the limitations of the different analyses are clearly explained. With this the limiting factors of the actual data around the interventions today in mHSPC are clearly shown.	-	No reply required.
3	6	As already mentioned in our remarks to the Scoping we do judge the integration of such a Benefit-Harm-Analysis as very interesting and as an important attempt to analyze and show the Balance between benefit and harm and in the end in a	-	No reply required.

		comprehensive way the actual benefit for the patient himself. The method used as well as the analyses themselves are presented in detail. Also here the limiting factors of the results are shown in a comprehensible way and rely on the limiting factors already explained in the chapter before (e.g. only few information about side effects). In spite of the limitations of these results here for the interventions in mHSCP the potential of such an analysis also for future HTA's is very well shown.		
4	7	The chapter gives a good and comprehensible overview of the studies considered.	-	No reply required.
5	8	The cost-effectiveness analysis can be reproduced. The choice of input parameters is supported. The results are plausible.	-	Thank you. No reply required.
6	9	In the context of the budget analysis, it is noted that the cost implications depend on the treatment strategy. The report states that it is very difficult to raise the treatment of patients with mHSPC. This gap should be closed.	It is recommended that the quantities of the corresponding drugs billed by the health insurance be used as a basis for estimating the treatment of patients with mHSPC.	This is an important point. We have also considered using billing data from health insurances for a better estimation. However, we encountered several issues when attempting to do so. First, ADT is given also in high-risk localized prostate cancer and there are various drugs in this category. Second, docetaxel is used across different cancer contexts. Third, enzalutamide and apalutamide may also be used in non-metastatic

				castration-resistant prostate cancer, and all drugs may be used in metastatic castration-resistant prostate cancer without concomitant ADT. Quantities of drugs billed would have to be cross-referenced with estimations of prevalence and incidence of mHSPC in Switzerland, which also bears some uncertainty (as outlined in the report). Thus, based on the data available in Switzerland, it is highly difficult to estimate the current market shares of the different treatment strategies in mHSPC. Attempting to do so would require a research project of its own and was thus outside of the scope of this HTA.
7	10	The chapter gives a good summary of the results of the study. The relevant findings are listed.	-	Thank you very much for your evaluation.
Specific Comments				

Stakeholder Comments

Full name	**Markus Ziegler, Heiner Sandmeier**
Function	
Organization	**Interpharma**

Comment number	Chapter	Comment	Suggested change	Response
General Comments				
1	General comment	The HTA report deviates from the common definition (Ludwig Boltzmann Institut. Methodenhandbuch für Health Technology Assessment Version 1.2012) which includes effectiveness/efficacy, safety and economic efficiency, as well as the evaluation of social, ethical, legal and organizational aspects. The important aspects social, ethical, legal and organizational are missing while a benefit-harm balance is introduced that is not part of common HTA.	Add social, ethical, legal, or organizational aspects that might be relevant in the context of this HTA.	Many thanks for your valuable comments. We agree that the common definition of HTA also includes an evaluation of ethical, social, legal and organizational aspects. However, this document is considered the scientific assessment report of our research group who is primarily responsible for conducting the presented assessments. The aforementioned aspects are commonly considered by the HTA appraisal team. This corresponds to the standard HTA process of the Swiss Medical Board.
2	General comment concerning benefit harm assessment	Assessing the benefit-harm balance is a quite uncommon aim of an HTA. There is no comprehensive description nor a validation of such a methodology. It is therefore premature	The benefit-harm assessment should be removed from the HTA due to the premature state of development and methodological weaknesses of the method.	We agree that quantitative BHA is not an official part of common HTA methodology. We also agree that a distinction needs to be made between regulatory decision-making and HTA. Regulatory agencies - such as

		to include a method that has not yet been sufficiently developed. The benefit-risk analysis is performed by regulatory bodies and not by HTA bodies. The benefit-risk assessment should remain within the responsible and competent authority. They base their decision on comprehensive documentation that is far deeper than the publications assessed by HTA. Your statement in response to a previous comment by a stakeholder: "Furthermore, BHAs are commonly conducted by the FDA and EMA for drug approval processes and it has become clear that BHA should be an integral part of HTA." is not correct. The comprehensive risk-benefit assessments by the responsible authority (FDA, EMA, Swissmedic etc.) cannot be equated with a non-validated experimental method for benefitharm balance for HTA. It raises issues, especially in cases, when a BHA assess a drug beyond the approved indication (e.g. in this HTA it is disregardes that approval of Abirateron acetat is for high-risk mHSPC only) and Docetaxel is not approved in mHSPC). Your statement: „... it has become clear that BHA should be an integral	In addition contradicting result may arise between the quantitative non-deleberative method of the BHA based on published clinical studies and the deliberative assessment/appraisal by Swissmedic (and other approval agencies) which are based on more comprehensive data. Such potential conflicts due to methodology need to be avoided.	*swissmedic* in Switzerland - are the competent authorities for benefit-risk assessment regarding market approval of individual drugs, following a clear deliberative process. However, BHA in a wider sense may also be used in other stages of the medical product life cycle, such as clinical development, HTA, post-approval monitoring, or supporting clinical practice, for example through clinical decisions. For example, the United States Preventive Services Task Force does include BHA in their assessment of the evidence and development of recommendations. We thus regard it as a flexible tool that can positively influence decision-making not just during market approval, but also in other evaluative contexts. The inclusion of a BHA was explicitly addressed in a board meeting of the Swiss Medical Board and a majority of its members approved the BHA as a potential innovation to HTA. A thorough assessment and balancing of the benefits and harms of a treatment is implicitly part of any HTA (as also stated in the general methods by LBI and IQWiG). While this assessment is commonly qualitative in nature, IQWiG for example keeps it at

		part of HTA." is in our view not backed by the references cited and to our knowledge not supported by regulatory agency nor is not part of common HTA.		their discretion to assess the benefit-harm balance for multiple outcomes by quantitatively aggregating them in a common metric (explicitly left unspecified in their methods). We argue that such a quantitative BHA provides additional value to HTA, since it allows a simultaneous consideration of benefits and harms while explicitly considering patient perspectives. Furthermore, BHA in the post-approval setting can be used to compare different drugs that have been approved for marketing (i.e., all having favourable benefit-risk profiles based to regulatory assessment). BHA may thus aid the comparative evaluation of multiple treatments, which is within the aims of HTA. In contrast to cost-effectiveness analyses (CEA), which consider costs and are typically context- and country-specific, BHA may be more comparable across contexts. In our analysis, we applied one of different established quantitative BHA approaches. We agree that more experience is needed with conducting BHA in the HTA setting. However, we do not agree that such research should not be conducted, since this is a precondition for gathering such experience. While some uncertainties

				exist, the applied method is a systematic and transparent approach that allows the exploration of its assumptions through various sensitivity analyses which we present in the report. The findings of the BHA are intended to complement the findings of the primary parts of the HTA, consisting of the evaluation of the clinical efficacy and safety, as well as the economic characteristics of the treatments. As also stated previously, the weight given to the findings of the BHA in the appraisal process is at the discretion of the SMB appraisal committee.
Specific Comments				
3	5.2.3.2	The conclude that there is a non-significant OS benefit in the low-volume subgroup for Apalutamide+ADT is not compatible with the final analysis of the TITAN trial.	Please update data from TITAN and other trials, discuss the maturity of clinical trial data used for the NMA as well as potential bias introduced by that in all the sectors of the HTA.	We are aware of the newly published data and updated our analyses regarding Overall Survival in the report. Please note that the new data was published on 29 April 2021, after submission of the second version of this report (v2.0, dated 23 April 2021) for stakeholder review.
4	6.1.2	The application of preference weights from other decision contexts to mHSPC patients is not an appropriate methodological approach. A benefit-harm assessment cited in moderate to severe COPD (Yu, 2014) and treatment with tamoxifen in	The benefit-harm assessment should be removed from the HTA due to the premature state of development and methodological weaknesses of the method.	We agree that the preference weights are a factor of uncertainty in the BHA. When applying preference weights from other contexts, certainly caution is warranted, and transferability needs to be assessed and critically discussed.

		female patients (Gail, 1999) concern totally different entities to mHSPC. Preference weights used in these cited study seem to be defined somehow arbitrary and not empirically assessed.		As outlined before, the published data did not allow for a detailed, individual AE-level BHA. We thus used generic, aggregate-level AEs in our analysis. By using "values [...] that have been applied in decision contexts", we evaluated the applicability of using these preference weights in the context of mHSPC and also compared them with the limited evidence available from preference studies in mHSPC. During this transferability assessment, we considered aspects of severity and consequences of AEs on an aggregate-level based on the CTCAE classification. As transparently stated in the report, the weights used were of generic nature. This approach was chosen by various studies using a similar BHA methodology. At an aggregate level, while considering severity and consequences, we deemed the weights to be broadly transferable. For example, the preference weights for a fatal outcome are expected to be similar, irrespective of whether it is due to prostate cancer or due to breast cancer (i.e., death is expected to be the most important outcome with a preference weight of 1.0). Similarly, we broadly assume transferability for

				mild to moderate and severe to very severe AEs. However, this assumption warrants further examination. In sensitivity analyses, we evaluated different preference weights for AEs, which did not result in a different conclusion. Within our understanding, even if a preference study was conducted, more accurate empirical preference weights would hence be unlikely to significantly change the benefit-harm balance. The future conduct of preference studies suited to inform BHA in the context of mHSPC may help to make better judgements about patient preferences and validate our findings. However, such a study was considered outside the scope of this HTA.
5	6.1.4	The statement: “This approach has been applied in benefit-harm assessments of various preventive and therapeutic treatments [51–55].” relates to studies including patients with COPD, primary prevention of cardiovascular disease and patients with blood pressure issues with the excpetion of (Gail, 1999, comment see above).	The benefit-harm assessment should be removed from the HTA due to the premature state of development and methodological weaknesses of the method.	The BHA approach used was originally developed in the context of breast cancer patients (Gail et al. 1999) and has subsequently been applied in a variety of other disease contexts. We agree that BHA approaches have to be carefully selected based on the context, which is also in line with recommendations of the EU IMI PROTECT project. However, we evaluated the used model as one of the most applicable

		The concept was not tested in critically ill patients.		and transparent given the aims of the analysis. An alternative approach that has frequently been used in cancer settings is Q-TWIST analysis. Such an analysis was not possible due to the unavailability of individual patient data. In addition, AEs of novel hormonal treatments are deemed not well captured with toxicity periods - a concept that relates better to chemotherapy. Additional testing and development of different BHA approaches in the cancer setting would indeed be valuable. We deem the used approach as most closely reflecting a clinical decision-making scenario (i.e., weighing expected survival gains against potential AEs of treatment), which can be applied well also in advanced cancer settings.
6	6.1.4	Two cutpoints are introduced: state following: above 60% for as net clinical benefit (or more clinical benefits than harms) and below 40% as net clinical harm (less benefits than harms). These cutpoint semm arbitrary and there is no reference to a validation?	The benefit-harm assessment should be removed from the HTA due to the premature state of development and methodological weaknesses of the method.	We chose 60% and 40% (instead of 50%, a threshold at equipoise where expected incremental net benefit is zero) since it is logical to be slightly risk averse when deciding on a more intensive treatment. That is, patients may opt for taking a treatment knowing they will achieve a certain incremental net benefit, and not if otherwise.

				While the choice of the cut-off at 60% or 50% can be debated, we deem it more plausible and pragmatic to take a higher probability to define a net clinical benefit than exactly the threshold of equipoise. The cut-off points of 40% and 60% were also applied in BHAs on aspirin, roflumilast, statins or inhaled corticosteroids (see references in report). We agree that a validation of these thresholds would make sense. However, the net benefits at a probability of around 50% (40% to 60%) were not ignored but considered an uncertainty zone in which decisions should be made with greater caution. Importantly, the ranking of treatments regarding their benefit-harm balance would not be influenced, regardless of the cut-off point.
7	6.2.1	We do not understand how an intervention without net clinical benefit can be cost-effective in the economic evaluation?	If not removed, the results of benefit-harm assessment would need a discussion in context with the results of cost-effectiveness assessment.	It is important to note that the BHA and the CEA had very different objectives (balancing treatment options for an individual patient independent of cost vs. judging the cost-effectiveness), perspectives (clinical decision-making vs. Swiss healthcare payer perspective), time horizons (2 years vs. 15 years), and model structure (event-based vs. time spent in health states; preference weights vs. utilities). The two analyses

				are thus not directly comparable and certain differences in results are to be expected. As such, they have to be interpreted as complementary to each other. We have now added this statement in the executive summary: *"[…] This analysis focused mainly on the evaluation of the benefit-harm balance from a clinical decision-making perspective, without taking any cost factors into account. Since the BHA pursues a different aim, its results thus should not be interpreted interchangeably but rather as complementary to the cost-effectiveness results."* The BHA found that docetaxel has a low probability of a net benefit over a 2-year time horizon from the perspective of the individual patient, due to the high rates of severe AEs. The CEA also showed that the QALYs gained with docetaxel compared to ADT alone over a 15-year time horizon were lower than with the novel hormonal therapies. The results of the two analyses are thus broadly consistent, since the comparison between the QALY benefit and the net clinical benefit is the most relevant for comparing results between CEA and BHA. Clearly, the much lower costs of

				docetaxel strongly influence cost-effectiveness in its favor, so that it is more cost-effective compared to the hormonal treatments from a Swiss healthcare payer perspective.
8	7.1.3	Articles reporting only costs or utilities were excluded in selection process. Why should publications reporting utilities be excluded?	Please justify exclusion of studies concerning utilities.	The literature review focused on studies reporting ICERs (i.e. costs per QALYs). Although studies reporting only costs or utility may provide relevant information, this restriction was necessary given the available budget.
9	7.1.3	The transferability of study results from countries with very different socioeconomic characteristics in Asia or South America, maybe also US is limited or not given.	Focus should be on C/E studies that have relevance to the Swiss setting.	We agree that adaptations of international ICERs to the Swiss setting should only consider analyses for jurisdictions with broadly similar socioeconomic characteristics as Switzerland (e.g. North, Central and Western European countries, the USA, Canada, Australia and New Zealand). In the present case, we decided to provide a broad overview of the existing literature. The main reason for not excluding articles from less similar jurisdictions (e.g. Asian or South American countries) was the fact that we also conducted a de novo CEA for Switzerland. The conclusions concerning the economic impact of the investigated treatments were based on our analyses and not on

				adapted estimations from other countries.
10	8.2.2.1	A willingness to pay (WTP) threshold of CHF 100,000 per QALY gained is used with the notion that it would sometimes be tentatively considered in analyses for Switzerland. However, there is no formal threshold in Switzerland. The Federal High Court (Bundesgericht) decided stated in its decision 9C_744/2018 from 1. April 2019 that it never had fixed a cost threshold. This is true also for a previous often cited decision on Myozyme (BGE 136 V 395) where it considered the opinion of some experts for a threshold of 100'000.- but in fact did not decide for fix a threshold. The results should therefore not be judged as cost-effective relating to this threshold which has no legal basis.	A threshold for statements of cost-effectiveness should not be used in the Swiss context; reporting of C/E ratios is sufficient.	Thank you for this comment. We are aware that there is no official threshold in Switzerland. This is why we only stated that the CHF 100,000 threshold is sometimes tentatively considered in CH. This is a neutral piece of information and does not advocate the use of a specific threshold. We re-formulated related sentences in the executive summary and in the conclusion of the report.
11	8.3.6.1 6.3.6.2	Medical resource use is based on numerous assumptions.	Please report whether the information given is in line with guideline recommendation.	Medical resource use is based on numerous assumptions as it incorporates many different parameters like e.g. drugs and their dosages, laboratory tests, physician visits, imaging resource use, and type of end-of-life (EoL) care. Drug dosages, imaging resource use and

				physician visits were discussed with our medical expert group.
12	C/E-analyses BIA	Treatment patterns are continuing to change over time and might substantially influence cost-effectiveness. In addition new drugs will change the distribution of market shares. Therefore, a long-term time horizon as applied in C/E analysis and budget analysis should be avoided.	A shorter time horizon such as 5 years should be applied to the base case analysis in costeffectiveness analysis and budget impact analysis.	We agree that many factors may influence the future treatment of mHSPC patients (new drugs, new scientific evidence, patents expiration). In the discussion we clearly stated that “the use of the investigated treatments for mHSPC is in constant evolution and may change very fast in the next few years”. Despite this limitation, which most cost-effectiveness analyses share, the time horizon of this CEA needs to be lifelong. The underlying methodological principle is that the time horizons of CE analyses need to be long enough to capture all relevant consequences of the initial strategy decision. E.g., NICE stipulates that the time horizon should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. As shown in Figure 19, we assume that almost all patients are dead after 15 years. We have hence chosen a time horizon of 15 years. Not all relevant costs and outcome differences would be captured after only 5 years.

				In two scenario analyses we have already investigated time horizons of 5 and 10 years (Table 43). Reducing the time horizon to 5 years had a big impact and resulted in higher ICERs.

References Stakeholder

Ludwig Boltzmann Institut. Methodenhandbuch für Health Technology Assessment Version 1.2012.) https://aihta.at/uploads/tableTool/UllCmsPage/gallery/Methodenhandbuch.pdf

Yu T, Fain K, Boyd CM, et al. Benefits and harms of roflumilast in moderate to severe COPD. Thorax 2014;69:616–22. doi:10.1136/thoraxjnl-2013-204155

Gail MH, Costantino JP, Bryant J, et al. Weighing the Risks and Benefits of Tamoxifen Treatment for Preventing Breast Cancer. JNCI J Natl Cancer Inst 1999;91:1829–46. doi:10.1093/jnci/91.21.1829

Stakeholder Comments

Full name	
Function	
Organization	**Janssen-Cilag AG**

Comment number	Chapter	Comment	Suggested change	Response
General Comments				
1	General comment	The Health Technology Assessment (HTA) Report evaluates the clinical effectiveness, safety, benefit-harm balance and health economic characteristics of docetaxel, abiraterone enzalutamide, and apalutamide in combination with ADT. A common definition / aim of an HTA-Report is to assess a health intervention according to the dimensions, effectiveness/efficacy, safety and economic efficiency, as well as the evaluation of social, ethical, legal and organizational aspects. (Ludwig Boltzmann Institut. Methodenhandbuch für Health Technology Assessment Version 1.2012.) Social, ethical, legal and organizational aspects have not been evaluated in this HTA-Report. Important aspects of a Health Technology Assessment are missing, therefore.	Add justification why the mentioned important aspects of a Health Technology Assessment were not addressed within the scope of the HTA. Describe social, ethical, legal, or organizational aspects that might be relevant in the context of this HTA. In case that no evidence or publications can be identified, that address these aspects, it would be helpful to describe potential aspects and discuss them in context with the results of clinical effectiveness, safety and health economics assessment.	Many thanks for your detailed and valuable comments. Please refer to our reply to comment number 1 by Interpharma.

2	General comment concerning benefit harm assessment	Assessing the benefit-harm balance is a quite uncommon aim of an HTA. The Methodenhandbuch für Health Technology Assessment Version 1.2012 published by the Ludwig Boltzmann Institute mentions benefit-harm balance three times in context with economic evaluations and quality adjusted life years (QALY). A comprehensive chapter that describes methods on how to conduct benefit-harm balance assessment is not provided. As stated in our stakeholder comment (18.02.2020) we believe that it is premature to include a method that has not yet been sufficiently developed, is validated or is standard of HTA. As stated in the supplement to the HTA Scoping Document from 2019-12-17, regulatory bodies are the responsible party for benefit-risk assessment. We agree and recommend to keep benefit-risk assessment within the responsible and competent authority, which bases their decision on comprehensive documentation, that is not in public domain. In your answer to our comment, you state: “Furthermore, BHAs are commonly conducted by the FDA and EMA for drug approval processes and it has become clear that BHA should be an integral part of HTA.” Original purpose of the approval process by FDA or EMA or Swissmedic is the assessment of benefit-risk-balance. In case that a drug was approved by FDA and/or EMA and/or Swissmedic, a competent authority has decided that the benefit-risk balance is sufficient for a	We strongly advise the removal of the benefit-harm assessment from the HTA due to potential conflicts in benefit-harm assessments between regulatory institutions and Swiss Medical Board. As will be pointed out later, we found a number of methodological weaknesses that support our advice to remove the benefit-harm assessment.	Please refer to our reply to comment number 2 by Interpharma.

		market authorization, based on comprehensive documentation. Conferring the judgement of a non-validated method for benefit-harm balance from the responsible authority (FDA, EMA, Swissmedic) to HTA, may raise issues, especially in cases, when a drug is not approved by FDA / EMA / Swissmedic, e.g. Docetaxel in mHSPC. Even more in cases where an assessment is performed in indications beyond the approved indication (approval of Abirateron acetat is for high-risk mHSPC only). The subordinate clause „... it has become clear that BHA should be an integral part of HTA." does not specify whose opinion is reflected. Is it the opinion of FDA/EMA/Swissmedic or the opinion of the authors of the HTA? Coming back to your scoping document (Version 2020-03-13) Part III, you state the following: Benefit-Harm Assessment: "Important initiatives by various HTA agencies and international collaboratives are currently ongoing aiming to integrate BHA in HTA processes, in an effort to make judgement about the balance of benefits and harms of novel treatments explicit and transparent [57,58]." Both citations (Mühlbacher, 1016; Ho, 2016) focus on the topic of incorporating patient preferences regarding benefits and risks into regulatory assessment. As Ho et al. is available in free full text we checked the text and the word HTA is not used in that document. Furthermore the authors state: "The project was used to develop a framework to help the Food and Drug Administration (FDA) and industry sponsors		

		understand how patient preferences regarding benefit and risk might be integrated into the review of innovative medical devices." From the cited literature we see no justification why benefit-risk assessment should be incorporated into HTA. In fact, the cited references confirm that the benefit-harm assessment is in the scope and responsibility of regulatory institutions not HTA agencies.		
Specific Comments				
3	12.2.1 12.3.1	Concerning transparency, for each search step of the literature review, the number of resulting hits should be reported, not just the final number.	Add information	We transparently reported the number of records retrieved through searches within each database in the appendix (12.2.1 and 12.3.1). Furthermore, we report the number of records across the literature search process, including search, screening and inclusion in Figures 1 and 17. Full reproducibility of the literature searches is possible based on the information provided in the report and its appendix.
4	5.1.1 5.2.1 12.2.1	The clinical study registry clinical trials.gov was searched, 50 hits reported, and these hits were further included in selection process as normal publications. The purpose of a study registry search normally is to identify ongoing or finished studies. On one hand you can assure that all studies are published, irrespective of their results. On the other hand, you can use information about ongoing studies to discuss if upcoming studies	Please report the results of study registry search and provide more detailed information about ongoing studies, that might have been published after your literature search or that might have an impact on the conclusions of your HTA in near future. Please discuss if any relevant publication is missing and	The records retrieved from clinical study registries were screened in their entirety as part of the systematic screening process. We did not consider trial registry entries to be publications, but as a record of ongoing or finished studies (as you also state). Thus, these records are part of the total 92 records considered eligible, of which we now added a complete list in the Appendix. Within these eligible records, we identified 14

		might have an impact on your HTA results and conclusions. Neither the first nor the second was done with the information you gained for study registry search.	potential hints for publication bias from study registry search could be found?	studies with data available from the 8 trials included in the analyses. Since some of the registry entries included data of relevance to this review (e.g., AE data), we kept these in the pool of records that could be used for data collection, in order to retrieve the most recent, complete and applicable data for each of the trials. We included a statement about ongoing studies in the discussion. Furthermore, we added a list of the eligible trials from which no data was available in the Appendix.
5	5.1.1 5.3.2	The authors report, that „recent published systematic reviews" were screened. As the systematic search in Medline and Embase was limited to controlled clinical trial or randomized trial, the only source for a **systematic** identification of reviews would be the database Cochrane Central. Further relevant databases like HTA/ DARE of the University of York (https://www.crd.york.ac.uk/CRDWeb/York) and/or INAHTA (inahta.org) were not considered. Methods for identification of systematic reviews or HTA are therefore not transparent. There is inconsistency between the reported methods and the results section. N=31 publications were excluded for reason "systematic review". If you were systematically searching / not systematically screening for	Please provide information about: • Databases used for identification of systematic reviews • whether you considered HTA or just systematic reviews • number of identified reviews/HTA and reasons for exclusion in full text • how "screening" and selection of systematic reviews were performed	The purpose of the systematic literature review in this HTA was to identify primary studies of relevance in mHSPC given the PICO. A systematic identification of systematic reviews was not considered within the scope of this HTA. However, the databases searched and search strategies used did allow for a systematic identification of systematic reviews on the subject. We identified systematic reviews in the title/abstract screening stage for our reference, but excluded them during full-text screening for primary studies. Given the aforementioned aims, the inclusion of further databases was not deemed of additional value and thus not included in the scope. We screened only the most recent

		systematic reviews, why did you exclude systematic reviews later? Systematic reviews were solely cited in chapter 5.3.2. As you did not report the number of identified systematic reviews / HTA, nor the selection process (inclusion/ exclusion criteria) it isn´t possible for the reader to judge if the cited reviews give a comprehensive overview or an unbiased discussion.	• in- and exclusion criteria • whether Internet sites of relevant HTA agencies have been searched Please discuss potential impact on reported results / conclusion.	systematic reviews for references and referred to these in the discussion of our findings, since several relevant studies were published within the 2 years prior to the start of the project. Thus, older reviews were considered to be no longer up to date. We clarified the wording in the methods section.
6	5.2.1 7.1.3	As a methodological standard for all publications that were excluded in full text the exclusion reason should be provided.	Please provide a list of publications that were selected in full text and give information about reason for exclusion for each.	We added the full list of records that were excluded from the review in the Appendix.
7	5.2.1 – flow-chart	The individual listing of excluded publication does not sum up to n= 133.	Have multiple responses been possible? Otherwise correct numbers.	Excluded records could fulfill multiple reasons for exclusion. We added this information in a Figure legend.
8	5.2.2	As identified by the authors, ENZAMET and LATITUDE trials are problematic for the NMA and should not be part of the base case.		This issue was discussed in-depth by the assessment team as well as with involved clinical experts. Since results from LATITUDE and ENZAMET constitute a relevant evidence base for decision-making in clinical practice and in order to ensure comparability with other meta-analyses in the context, we decided to include these studies in the base case. However, we transparently highlight the uncertainty of using this evidence in the GRADE assessment, sensitivity analyses and discussion of our findings.

9	6.1.1	We assume, in regard to the cited references in the scoping document (57, 58 ibid.)(see comment Nr. 2 above) that the authors of the HTA wanted to conduct a benefit-harm assessment and to incorporate patient preferences regarding benefits and risks. Therefore, the use of specific AEs in the BHA would be much more relevant and a weighting according patient preferences.	See comment Nr. 2	We agree that the use of specific adverse effects (AEs) in benefit-harm assessment could provide additional detail that may be relevant for making treatment recommendations. As part of this HTA, we planned to conduct such an analysis. However, there were two major issues that precluded us from doing so: First, detailed and specific AE data was unavailable for most of the trials, which would be a precondition for a more granular analysis. We strongly encourage the full and comprehensive reporting of adverse effects data for industry-sponsored trials. Data provided for LATITUDE and ENZAMET are exemplary in this respect. Second, patient preference values suitable for quantitative BHA were generally unavailable for most AE outcomes. Thus, generic values had to be used for the present assessment. Overall AEs categorized by the CTCAE classification were reported by all trials. The CTCAE classification clearly defines AEs in terms of their severity, duration and consequence. Due to this standardization, aggregated AEs at different severity levels were considered to correspond well across individual AEs and thus likely to yield preferences of broadly

				similar magnitude on an aggregate level. In light of the available evidence, we thus considered the use of aggregated AEs and generic preference weights the most appropriate approach. Further preference research in the context of mHSPC would be necessary to provide preference weights for BHA.
10	5.2.3.2	The Authors conclude that there is a non-significant OS benefit in the low-volume subgroup for Apalutamide+ADT. Please note that the final analysis of the TITAN trial has been published, which shows otherwise.	Please discuss maturity of clinical trial data used for the NMA as well as updates from other trials as well as potential bias introduced by that in all the sectors of the HTA (see comment 4).	Many thanks for raising our attention to the updated results. We are aware of the newly published data and updated our analyses regarding Overall Survival in the report. Please note that the new data was published on 29 April 2021, after submission of the second version of this report (v2.0, dated 23 April 2021) for stakeholder review.
11	6.1.2	The authors state: “The preference weights represent the relative importance (or seriousness) of the outcomes, where a higher value means that patients have a stronger preference to avoid the outcome. Empirically determined preference weights, for example by preference-elicitation surveys, were not available for the specified outcomes. We therefore used generic values on a scale of 0 to 1.0 that have been applied in other decision contexts [50,51]…”. We don´t agree that applying preference weights from other decision contexts to mHSPC patients is an appropriate methodological approach. One	See comment Nr. 2	We agree that the preference weights are a factor of uncertainty in the BHA. When applying preference weights from other contexts, caution is certainly warranted, and transferability needs to be assessed. As outlined before, the published data did not allow for a detailed, individual AE-level BHA. We thus used generic, aggregate-level AEs in our analysis. We evaluated the applicability of using preference weights applied in other decision contexts in the context of mHSPC by comparing them with the

		cited reference reports a benefit-harm assessment in moderate to severe COPD (Yu, 2014), a totally different entity to mHSPC. Preference weights used in the cited study seem to be defined somehow arbitrary and not empirically assessed. The second reference was published 1999 (Gail, 1999), reflecting treatment with tamoxifen in female patients. Here again, we couldn´t find an empirically evaluation of patient preferences.		limited evidence available from preference studies in mHSPC. In this transferability assessment, we considered the severity and consequences of AEs based on the CTCAE classification, and compared them with preference values in studies regarding harm outcomes with a similar expected impact on patients. As transparently stated in the report, the weights used were of generic nature. This approach was chosen by various studies using a similar BHA methodology. At an aggregate level, while considering severity and consequences, we deemed the weights to be broadly transferable. For example, the preference weights for a fatal outcome are expected to be similar whether it is fatality due to prostate cancer or due to breast cancer (i.e., death is expected to be the most important outcome with a preference weight of 1.0). Similarly, we broadly assume transferability for mild to moderate and severe to very severe AEs. However, this assumption warrants further examination. In sensitivity analyses, we evaluated different preference weights for AEs, which did not result in a different conclusion. Within our understanding, even if a

				preference study was conducted, more accurate empirical preference weights are hence unlikely to significantly change the benefit-harm balance. The conduct of preference studies suited to inform BHA in the context of mHSPC may help to make better judgements about patient preferences and validate our findings. However, such a study was outside the scope of this HTA.
12	6.1.4	The authors state the following: “This approach has been applied in benefit-harm assessments of various preventive and therapeutic treatments [51–55].” Only one of the cited studies applied the concept to breast cancer patients (see limitations to this study in comment 9, Gail, 1999). All remaining studies included patients with COPD, primary prevention of cardiovascular disease and patients with blood pressure issues. The concept was not tested in critically ill patients.	See comment Nr. 2	Please refer to our reply to comment number 5 by Interpharma.
13	6.1.4	The authors state following: “A probability above 60% was interpreted as net clinical benefit (or more clinical benefits than harms), below 40% were interpreted as net clinical harm (less benefits than harms), or as neither harmful nor beneficial between 40% and 60%.” Are these thresholds validated? Could you provide a valid reference for these cut-off-points.	See comment Nr. 2	Please refer to our reply to comment number 6 by Interpharma.

14	6.2.1	The authors state following: “The corresponding probability of net benefit were 97.5%, 83.3%, 89.4%, respectively. However, we did not find a net clinical benefit for ADT + docetaxel, which had a lower probability to experience a net clinical benefit (14.9%).” We don´t understand why an intervention without net clinical benefit can be cost-effective in the economic evaluation?	Please discuss the results of benefit-harm assessment in context with the results of cost-effectiveness assessment.	Please refer to our reply to comment number 7 by Interpharma.
15	7.1.2	Further relevant databases like HTA/ DARE /NHS-NEED of the University of York (https://www.crd.york.ac.uk/CRDWeb/York) and/or INAHTA (inahta.org) were not considered in the search.	Please discuss potential bias to your HTA	According to the website (https://www.crd.york.ac.uk/CRDWeb/), “Bibliographic records were published on DARE and NHS EED until 31st March 2015. Searches of MEDLINE, Embase, CINAHL, PsycINFO and PubMed were continued until the end of the 2014.” Considering that the DARE and NHS EED are no longer updated while most of the evidence in this HTA was published after 2014, and considering that the CRD databases (including DARE and NHS EED) are included in the Cochrane Library, a specific search in this database was not considered. Of note, the rebuild of the new platform by the INAHTA was launched in June 2020, after the conduct of the searches for this HTA.
16	7.1.2	The authors state following: “… and published search strings for health economic analyses of the NHS EED [64].”	Please clarify if the NS EED Database was searched and add where appropriate results	The Cochrane Library is maintaining versions of the Database of Abstracts of Reviews of Effects (DARE) and the

			in section 12.3.1 and flow-chart.	NHS Economic Evaluation Database (NHSEED) until at least 2021. Therefore, economic analyses of the NHS EED were automatically included through the Cochrane Library search.
17	7.1.3	Articles reporting only costs or utilities were excluded in selection process. Why should publications reporting utilities be excluded?	Please justify exclusion of studies concerning utilities.	Please refer to our reply to comment number 8 by Interpharma.
18	7.1.3	Methodological standards for HTA recommend that selection of literature should be performed by two reviewers, but was done here by a single reviewer.	Please discuss resulting uncertainty in discussion section.	We agree. This is especially true for the clinical assessment (due to the fact that the number of hits is usually high, and the title of the articles may sometimes be misleading). In this case, the number of hits in the economic search was relatively low. In addition, the screen for CEA is usually simple since all relevant articles usually mention CE, QALY, or ICERs in their titles/abstract. The screen with a single reviewer was judged sufficient.
19	7.1.3	The limited number of publications with similar socioeconomic characteristics as Switzerland should not be used to justify the inclusion of publications with less similar jurisdictions. The transferability of study results from Asia or South America, maybe also US is limited or not given.	We recommend focusing on C/E studies that have relevance to the Swiss setting to get a more precise view of the current evidence concerning C/E. Although C/E studies with low scoring on CHEERS checklist should be ignored.	Please refer to our reply to comment number 9 by Interpharma.

20	7.2.3	The quality of reporting of C/E studies was judged by the CHEERS checklist and quality was described in an aggregated way. Reporting of individual judgement of studies would be interesting and transparent.	Please provide full details to the CHEERS checklist of each C/E study.	We provided in the appendix the study evaluations according to the CHEERS checklist.
21	8.2.2.1	The Markov cohort simulation consists of three states. It remains unclear in what state the patient cohort starts simulation. Is a decision tree part upstream that covers treatment phase?	Please clarify distribution of patients in health states at the beginning of simulation.	Thank you for this valuable comment. We clarified the distribution of patients at model start as follows in 8.2.2.1: *All patients started in the progression-free disease state and could either stay in this state, progress or die. Patients were distributed between the three states by using parameterized survival curves estimated from clinical trial results. Patients progressing under mHSPC treatment entered CRPC,…*
22	8.2.2.1	As the authors state: "…willingness to pay (WTP) threshold of CHF 100,000 per QALY gained, which is sometimes tentatively considered in analyses for Switzerland." There is no formal threshold and results should therefore not be judged as cost-effective relating to this threshold. It would be more helpful to compare to c/e ratios from other indications /indication groups.	Just report C/E ratios and do not use a threshold for statements of cost-effectiveness.	Please refer to our reply to comment number 10 by Interpharma.
23	Table 23	References are used that were not identified in the systematic literature search, e.g. Heijnsdiijk, 2012 or Davis, 2015.	Please explain the approach to identify additional literature and how you can assure that all relevant publication were identified.	Since the utility results of the HE literature search did not result in completely consistent utility findings for each state and treatment, we coupled the results of the HE literature search with hand-searches as pre-defined at

				the beginning of the HE analysis (Chapter 8.1 last bullet point): Additional targeted searches, complemented with hand-searches of grey literature and the World Wide Web (non-systematic searches), were performed in order to identify health resource use and costs that were not available from the above-mentioned sources.
24	Table 23	Further utility values used show a relevant underlying uncertainty as they are based on assumption, low quality studies or are based on study that do not reflect the mHSPC patients (see comment 22)	Due to the uncertainties related to utility values the validity of cost-effectiveness analysis is low. Due to the low validity of the C/E analysis statements concerning cost-effectiveness by using a threshold should be avoided.	Utilities for the progression-free state of our model correspond to the utilities used in the CEA by Sathianathen et al. 2019 (eligible study found in HE literature search, please see 7.2.1). The selected utilities were also regarded as sensible by the medical experts and are in line with those in the other publications quoted (out of which several were again found eligible in the HE literature review). For the post-progression state, we distinguished between a post-progression treatment state and terminal illness, in contrast to Sathianathen et al. He based his CRPC single state utility partly on Heijinsdijk (his input table 1). We used the original Heijinsdijk terminal illness utility of 0.40 for men with prostate cancer, and a higher utility of 0.635 for

				further line treatment (based on Aguiar et al. 2019, eligible study). In general, utility values are afflicted with uncertainty, especially if they cannot directly be derived from patient level preference data from one relevant RCT directly comparing all treatments under investigation. We used the best evidence available from the official HE literature search, coupled with additional hand searches and discussions with the experts. Most importantly, we performed univariate sensitivity analyses for the utilities and AE disutilities and showed that “variation in utilities had a high impact on the ICERs but did not change the overall conclusion”. Last but not least utilities were reported as a limitation to our model in the discussion section mentioning that further trials comparing treatment strategies directly and collecting utility data from the patients would be beneficial.
25	Table 23	Davies, 2015 considered utilities of patients with secondary hyperparathyroidism and chronic kidney disease requiring dialysis. This is a specific population and utility values can´t be transferred to oncological patients. The study of Davies, 2015 includes male (45,2%) and female patients (54,8%) and doesn´t report	Please argue and justify why you think that utilities of patients with secondary hyperparathyroidism and chronic kidney disease requiring dialysis are transferable to patients with	Not all adverse event disutilities used in our CEA were based on Davies 2015. E.g. the highest utility decrement (-0.37) for febrile neutropenia could also be taken from Sathianathen 2019 (as the health state utilities). There is sparsity of utility decrements for the

		gender specific values. Due to gender-differences the utility values can´t be transferred to a pure male population. The value of -0.09 as utility decrement for fracture can´t be verified in the publication of Davies, 2015. We assume that the value was calculated as mean of the values of hip fracture and arm fracture. Calculations like this should be reported transparently so that the reader of the report can realize underlying uncertainties. Ischaemic heart disease and cerebrovascular disease are not mentioned like this in the publication of Davies, 2015. Corresponding values indicate the authors used the values for myocardial infarction and stroke.	mHSPC and how you considered gender differences. Please report all calculations that changed data from original publication. Please state if the definition of ischaemic heart disease and myocardial infarction is identical in your work and the publication of Davies, 2015. As well as for cerebrovascular disease and stroke and justify why you think that value from the publication of Davies, 2015 can be used for patients with mHSPC.	other important grade 3 to 4 AEs which were considered important based on the Swiss medical experts and leading to hospitalizations. For this reason, we used as an approximation the utilities cited in Davies as the best available estimates. From a medical point of view, we regarded them as approximately sensible for patients with mHSPC. In order to investigate uncertainty attached to these disutilities, we performed deterministic and probabilistic sensitivity analyses. These did not change the overall conclusion of the analysis. Since the definitions of major clinical events stated in Davis did not always 100% match our selected AEs, we discussed comparability of listed AEs. We considered disutilities for cardiac failure and heart failure as very similar, as well as disutilites for ischaemic heart disease and myocardial infarction. We better clarified these assumptions in Table 23 of the SMB report. Yes, the utility decrement value of -0.09 for fractures was based on the mean of the disutilities for arm and hip fractures from Davies 2015. We clarified this point as well in Table 23 of the SMB report.

26	8.1	The authors state following: We investigated the cost-effectiveness of ADT in combination with docetaxel, abiraterone, enzalutamide, or apalutamide, compared with each other and to ADT alone." No pairwise comparison to ADT alone is reported.	Please provide information about pairwise comparison of of ADT in combination with docetaxel, abiraterone, enzalutamide, or apalutamide to ADT alone.	We included an additional table (Table 39) in the report which directly shows the increments vs ADT (so far results can be derived from Table 37).
27	8.2.2.1	The authors developed a 3-state Markov cohort simulation model. NICE DSU TSD 19 have stated that a simple PSM should at least be presented alongside another approach for the purpose of validation.	Please discuss potential implications and bias.	It was not in the scope of this project to construct a second independent model for validation purposes. We preferred a Markov cohort model over the alternative option of a partitioned survival model, due to the latter's assumption of structural independence between endpoints and to have a clear structural link between endpoints and overall survival. As each transition is explicitly modelled in a Markov model, treatment effects, other covariates and alternative scenarios can impact specific components of the simulation. Also, the Markov model allowed us to consider time in health state by the use of tunnel states.
28	8.3.2	The authors conducted a meta-analysis of OS and PFS KM survival data for patients treated with ADT across all included studies.	Definitions of PFS across studies is not consistent. Compared to TITAN trial, only ARCHES shares a similar definition of disease progression (rPFS) and ADT definition. Incorporating other trials here and creating a	We agree that different PFS definitions have been used in the trials selected for our analyses. As you correctly state, PFS definitions vary across studies and not all measured PFS endpoints were reported in all studies. In order to perform a meta-analysis of similar and comparable PFS endpoints

			pooled ADT PFS arm is resulting in a longer time spent on treatment than if only TITAN (+ARCHES) would be used. This will increase the treatment costs pre-progression, but will also result in a shorter time spent post-progression after receiving ADT alone (so the cost of subsequent treatments after ADT will be underestimated also). We do not agree that a pooled arm should be used here for PFS, but instead applying parametric distributions to the TITAN ADT arm would be more appropriate.	under ADT, we selected only the PFS endpoints we deemed most closely reflecting a meaningful clinical progression (either by leading to a decrease in HRQoL, an increase in symptoms or a change in treatment), as defined in the scoping document. These were **clinical progression-free survival (cPFS**; time to progression in clinical symptoms or radiographic findings, or death) **or radiographic progression-free survival (rPFS;** time to progression in radiographic findings, or death). The same procedure was applied for the base case of the NMA of the clinical part, for which the evidence selection is transparently reported in the Appendix. For example, the STAMPEDE Abiraterone arm was only included for OS but not for PFS because only FFS was available (Please see Table 22).
29	8.3.2	The rationale for using a piecewise curve after the KMs is unclear here, since at some point it seems that patients will die pre-progression (which is not expected in mHSPC). It would be more appropriate to apply standard parametric distributions to predict long-term outcomes rather than using a piecewise approach unless there is a clear reason for doing so.		We used the published KM-estimates as long as they were available from the trials and appended them with parametric distributions (for the extrapolated parts). Of course, we could have applied the parametric distributions from the start, but this would had only been an approximation of the KM estimates. We preferred the first approach as the values were closer to the real values.

30	8.3.6.1 6.3.6.2	Medical resource use is based on numerous assumptions.	Please report whether the information given is in line with guideline recommendation.	Please refer to our reply to comment number 11 by Interpharma.
31	8.3.7.1	Were compliance rates taken into account in the model of Drug and administration costs as would be reflective of clinical practice?		We assumed 100% compliance in the model. We added this point as a model limitation in chapter 8.6 of the report.
32	C/E-analyses BIA	A long-term time horizon as applied in C/E analysis and budget analysis should be avoided for several reasons: • treatment patterns / settings might change over time and might have a substantial influence on cost-effectiveness • new drugs will be approved and will lead to a changing distribution of market shares	Base case analysis in cost-effectiveness analysis should consider a shorter time horizon. Budget impact analysis should cover a shorter (5 years) horizon.	Please refer to our reply to comment number 12 by Interpharma.

References Stakeholder

Ludwig Boltzmann Institut. Methodenhandbuch für Health Technology Assessment Version 1.2012.) https://aihta.at/uploads/tableTool/UllCmsPage/gallery/Methodenhandbuch.pdf

Mühlbacher AC, Juhnke C, Beyer AR, et al. Patient-Focused Benefit-Risk Analysis to Inform Regulatory Decisions: The European Union Perspective. Value Health J Int Soc Pharmacoeconomics Outcomes Res 2016;19:734–40. doi:10.1016/j.jval.2016.04.006

Ho M, Saha A, McCleary KK, et al. A Framework for Incorporating Patient Preferences Regarding Benefits and Risks into Regulatory Assessment of Medical Technologies. Value Health 2016;19:746–50. doi:10.1016/j.jval.2016.02.019

Yu T, Fain K, Boyd CM, et al. Benefits and harms of roflumilast in moderate to severe COPD. Thorax 2014;69:616–22. doi:10.1136/thoraxjnl-2013-204155

Gail MH, Costantino JP, Bryant J, et al. Weighing the Risks and Benefits of Tamoxifen Treatment for Preventing Breast Cancer. JNCI J Natl Cancer Inst 1999;91:1829–46. doi:10.1093/jnci/91.21.1829

Stakeholder Comments

Full name	**George Thalmann**
Function	Präsident
Organization	**Swiss Urology**

Please note that the comments were based on first report version (v1.0 submitted to SMB on February 26, 2021).

Comment number	Chapter	Comment	Suggested change	Response
General Comments				
1	Excecutive Summary Introduction	We should state until when studies were included.		Thank you very much for your valuable feedback. We stated the date of the database searches (February 2020) in the methods section of the executive summary. We added the search date for conference proceedings (April 2020) to this section.
2	Budget Impact Analysis Seite 13	Seit die medizinischen Onkologen sich zunehmend um die ADT kümmern sind Ansätze, wie die intermittierende ADT gestorben. Die EORTC Studie 30891 hat gezeigt, dass wenn man erst bei Beschwerden die ADT einleitet gut 30% nie eine ADT gebraucht haben. Daher vergleichen wir hier ADT vs ADT plus irgendetwas anderes. Aber die Morbidität der ADT wird nicht miterfasst.		You raise the interesting point that ADT may not be beneficial in all patients with newly diagnosed mHSPC, but that deferred treatment with ADT may also be an option (according to EORTC 30891). Since the individual characteristics and patient preferences play an important role in prostate cancer care, this may indeed be an option to discuss with patients and of relevance in clinical practice in Switzerland. However, this

		Die EORTC 30891 Studie kommt zum Schluss: "Immediate ADT resulted in a modest but statistically significant increase in overall survival but no significant difference in prostate cancer mortality or symptom free survival. This must be weighed against the side-effects of lifelong androgen deprivation on an individual basis with the option of deferred treatment in a substantial number of patients." Wer wirklich von einer sofortigen Therapie profitiert sind Patienten mit PSA über 50 und einer PSA-Verdopplungszeit unter 6 Monaten.		HTA focused on a comparison of different (active) treatments in mHSPC. Since the estimation of the frequency of different treatment strategies was barely possible based on available data from Switzerland, we assumed that all patients receive active treatment in the Budget Impact Analysis. However, we added a statement regarding deferred ADT to the overall discussion.
3	Results & Discussion Assessment of Clinical Effectiveness and Safety Seite 15	"In summary"… We need to bear in mind that we do not know whether immediate or deferred combination therapy makes a difference. This with the health economic calculations in mind.		We agree that this was not part of the evaluation in this HTA.
4	Cost-Effectiveness Analysis Seite 17, 1. Abschnitt	Hier machen wir einen Überlegungsfehler, wenn wir 15 Jahre nehmen. Nach Einleitung einer ADT erfolgt nach 18-36 (STAMPEDE 48) Monaten eine Tumorprogression. Anschliessend kommen dann andere Behandlungen zu Anwendung. Die wenigsten überleben so lange.		CEA haben normalerweise einen lebenslangen Zeithorizont. Der Zeithorizont sollte lange genug sein, um alle Unterschiede in den Kosten und QALYs bzgl. der zu vergleichenden Therapien auch über die Behandlungsdauer hinaus zu reflektieren. Ein Zeithorizont von 15 Jahren bedeutet dabei nicht, dass wir annehmen, dass alle Patienten 15

				Jahre überleben. Wir beobachten alle Patienten maximal 15 Jahre und berücksichtigen dabei soweit möglich auch die Folgetherapien. Einige Patienten werden natürlich früher sterben.
5	Budget Impact Analysis Seite 18, 1. Abschnitt	Das Problem mit docetaxel ist, dass es eine Chemotherapie ist. Viele Patienten wollen diese nicht und bekommen aus diesem Grunde dann abi, enza oder api.		We agree that patient preferences for treatments play an important role in clinical practice. In addition, the market share among mHSPC patients in Switzerland is unknown. Enzalutamide and apalutamide reached market approval relatively recently and it is reasonable to assume that their market share in mHSPC will change in the next few years. Due to the uncertainties, we provided several scenarios based on various assumptions about market shares.
6	Benefit-Harm Assessment Seite 18, 2. Abschnitt	Viele ältere Patienten in diesem alten Patientengut werden nicht mit docetaxel behandelt genau wegen den Nebenwirkungen.		This is in line with the feedback that we have also received from other experts.
7	3. Abschnitt	I am not sure I would swallow that stuff with these numbers.		No reply required.
8	5.2.4.2 Systemic mHSPC Treatment ADT + abiraterone vs. ADT alone	This is always a problem. We do not have a cross-over for ADT alone. ADT alone still has the option of treatment with abi and could potentially have a better or the same outcome with a deferred approach. This is never considered in all these trials. From a		We agree that deferred treatment is a treatment option and that consecutive treatments after progression were not considered in intention-to-treat analyses of the trials. Participants in intervention and comparator groups may have received different post-

	Seite 46	survival and cost perspective this would be important.		progression treatments, which could also have had an effect on overall survival and quality of life. While an important consideration, deferred treatment strategies were not within the scope of this HTA and the data available did not allow us to make conclusive statements in this respect.
9	ADT + enzalutamide vs. ADT (± nsAA) Seite 47, letzter Abschntt	Be aware that the EORTC QLQ-C30 was developed for chemotherapy and does not well reflect outcomes in patients with ADT and second line hormonal therapy, thus a poorer QoL in the chemotherapy arm is to be expected.		We agree that the EORTC QLQ-C30 captures some domains of quality of life that overlap with the typical side effect profile of chemotherapy (e.g. nausea, vomiting, diarrhea), thus potentially leading to poorer quality of life estimates for patients receiving chemotherapy. We added a statement to the discussion in the assessment of clinical efficacy and safety.
10	Table [27]: CRPC first line and CRPC second line treatment durations. Seite 96	All these series are highly heterogeneous mixed pickles		We agree that the treatment context of metastatic castration-resistant prostate cancer (mCRPC) is highly heterogenous and strongly depends on prior systemic and local treatments. We decided to make assumptions in the model to estimate the drugs of two further line treatments when patients have reached a castration resistant prostate cancer state, their proportional use, and their duration when patients have achieved a castration resistant prostate cancer

				state. The use of further line treatments was discussed with the medical experts and is in line with the literature. As for the further line treatments themselves, estimates were also discussed with the experts. While further line treatment can of course vary from patient to patient, we needed to make reasonable and sometimes simplifying assumptions for modelling purposes. It is not possible to model too complex combinations of a large variety of further line treatments. We clearly stated that these assumptions correspond to a simplified approach and outlined it in chapter 8.3.6.2.
11	Table 22 [new Table 29]: Imaging resource use. Seite 98	PSMA PET? PSMA PET is now accepted for initial staging in prostate cancer and will certainly have an impact on treatment and treatment outcome. Earlier diagnosis might “positively” influence survival data.		We agree that the introduction of PSMA-PET may have a relevant impact on treatment and outcomes of patients with mHSPC. While it may be used instead of conventional CT for initial staging, we understood from our conversations with experts that this has not yet been implemented as a standard of care. Thus, we used estimates for conventional CT in our economic model. The use of PSMA-PET for initial staging would increase the overall costs of the different strategies, but we judged this to be non-differential between treatments

				and thus unlikely to have an impact on the overall assessment which would change our conclusions.
12	7.3.7.5 End-of life care costs Seite 100	"We adjusted this cost by the percentage of EoL prostate cancer patients in Switzerland who are hospitalised at the EoL (which we assumed to be 65%), resulting in one-off EoL costs of CHF 22,816*0.65 = CHF 14,830 per patient." Difficult to understand for a non-economist.		EoL costs are difficult to estimate and approximate. We based our estimate for EoL costs on costs previously observed for end of life at the Cantonal Hospital in Lucerne. Since we assumed that not all, but only a certain percentage of patients, would receive EoL treatment in hospital (some might receive EoL care at home), we multiplied the EoL costs with assumed hospitalization in 65% of patients. This has been done in a similar way in another published analysis (*e.g. Panje et al. 2020; see reference below*).
13	Figure 17 [new Figure 19]: Pooled overall survival ADT graph and all individual Kaplan Meier graphs. Seite 103	ADT Alone?		Yes, Figure 17 (Figure 19 in updated report) shows OS graphs for ADT monotherapy only (curves from different included studies as well as a pooled ADT curve based on a meta-analysis).
14	Figure 20 [new figure 22]: Base case efficiency frontier,	Docetaxel the patent protection is over. So it is cheaper now. Patent protection still functional for abi, enza and api		The figure title had already been changed in the latest version. The x-axis now shows QALYs. For docetaxel, the cheapest available price for the drug on the "Schweizer

	Note that effectiveness is measured in QALYs (it could also be measured in Lys) So why don't you write QALYs? Seite 108	Do I understand correctly: in case of enza for >6 times the cost effectiveness does not even double?		Spezialitätenliste" was used when the analysis was performed, corresponding to e.g. the price of Docetaxel Accord®. We applied the following approach: If original drug and generics/biosimilars were listed on the Swiss specialty list we consistently applied the lowest price for all treatments in our model. Double in relation to and for what? The cost-effectiveness ratio of ADT + abiraterone is CHF 294,163, whereas the ratio increases to more than CHF 1 million for ADT + enzalutamide.
15	7.6 Summary, discussion and conclusion Seite 118	Surgical castration is the best ADT and costs once maybe 2'000.- CHF. Has been forgotten, as often patients do not like the idea and industry has been pushing medical castration.		Thank you for this comment. We discussed the surgical castration at the beginning of the project. Expert feedback estimated the percentage of patients with an orchiectomy to be smaller than 5%. Given this low percentage we did not cover patients undergoing orchiectomy.
Specific Comments				
<u>References Assessment Team</u> Panje et al. 2020: A cost-effectiveness analysis of consolidation immunotherapy with durvalumab in stage III NSCLC responding to definitive radiochemotherapy in Switzerland; https://www.sciencedirect.com/science/article/pii/S0923753420359287				