Subtitles and transcriptions

Subtitles and transcriptions are available for selected materials for purpose of helping users understand the contents of the educational sessions.

Uncertain words have been indicated with ?? before and after the part. Parts that could not be understood at all have been indicated as [Audio Not Clear].

Every effort has been made to faithfully reproduce the audio of the sessions as recorded. However, no responsibility is accepted for mistakes or omissions. ESO does not endorse any opinions expressed in the presentations.

How to do a critical appraisal of scientific papers

Dr Seruga: Hello everybody. Good evening. Welcome to this online session of the European School of Oncology. The topic of the today's presentation is "How to do a critical appraisal of scientific papers." I'm medical oncologist at an Institute of Oncology, Ljubljana, and discussing will be doctor Templeton, who is medical oncologist at St. Claraspital, Basel in Switzerland. So, our critical thinking can be compromised for several reasons. One of the main reasons is for sure lack of knowledge and necessary skills, which we do not develop during our training. And the second reason may be exhaustion. When we are tired and we, medical oncologists and oncologist overall, are tired during our practise. And we can take shortcuts for our decisions, which means that we simply prescribe what's approved and we don't evaluate the risk and benefits of drugs we prescribe. And then, the last but not least, other may influence our thinking. And this, even if we have knowledge and skills, and this can be even more pronounced if we do not have necessary skills and if we are tired. Conflict of interest is a clash between requirements, which in our case is to care for patients, and interests, which can be financial or intrinsic, personal. So, we know that payments from pharmaceutical industry influence or prescribing practise. We know that experts on guideline committees have substantial financial conflicts of interest. And we know that declaring conflict of interest may not prevent bias. We oncologists are very often very driven people and we are engaged in publication and research to achieve career advancements. We may want to receive accolades from peers and professional societies, and we can be competitive for grant funding. So, why to use toxic and expensive cancer drugs? There are only two aims. One is to improve quality of life of cancer patients and/or to improve overall survival. So, these are patient-centred outcomes. In other words, relevant endpoints. Unfortunately, cancer trials do not necessarily measure these outcomes and surrogate measures or endpoints are not a direct measure of clinical benefit. They are direct measures if they are properly validated as such. Significant P-value usually defines positive randomised clinical trials. But what is P-value? The P-value is a number calculated from a statistical test that describes how likely you are to have found a particular set of observations if the null hypothesis were true, the P-value is unfortunately a very fragile thing. And we often misunderstand and misuse it. American Statistical Association gave a few statements about P-value. So, P-value does not measure the probability that a hypothesis is true, or the probability that the data were produced by the chance alone. Scientific and policy decisions should not be based only on the P-value less than 0.05. And the P-value, or statistical significance, does not measure a size of an effect or the importance of results. By itself, P-value does not provide a good measure of evidence regarding the model or hypothesis. So, what should I ask myself when I question positive randomised clinical trials? What is the endpoint? What is the effect size? What is tolerability? And how do all this matter to patients in everyday clinical practise? Randomised clinical trials in oncology evolved over the last few decades. They became larger and contemporary randomised clinical trials largely measure putative surrogate endpoints, such as progression-free survival or time to progression. And they're almost exclusively funded by the pharmaceutical industry. So, what is progression-free survival? Disease progression was intended for use in clinical trials to screen new drugs. It was intended to describe what happens to tumour during the therapy, not to infer a meaningful benefit to a patient from those changes. So, what improvement in PFS and other similar endpoints means for patients? Is their prognosis or

overall survival improved? And the other relevant question is, is their health-related quality of life improved too? We have evidence that the strength of association between surrogate endpoints and overall survival in oncology is weak. Prasad and colleagues conducted a systematic review of trial-level meta-analysis, which evaluative of benefit in progression-free survival is correlated with the benefit in overall survival. So, they identified 36 articles in which 65 specific correlations between a surrogate endpoint and survival were analysed. More than half of correlations were of low strength, a quarter were of medium strength, and only 23% of these correlations were highly correlated. Unfortunately, all validation studies used only a subset of available trials. You are encouraged to send some questions. So, the relevant questions are why surrogate endpoint may not predict outcomes in clinically relevant endpoints, such as overall survival? The reason may be technical factors. So, PFS is not measured appropriately, accurately due to more or less obvious reasons, such as measurement bias, evaluation bias, attrition bias, informative censoring. In a moment, I will discuss informative censoring in more details. And then the second reason may be that surrogate endpoint simply does not predict overall survival. It may not have a causal role in the clinically relevant endpoint, such as for example, pathologic complete response, or drug may change relationship between surrogate endpoint and overall survival. For example, there may be off-target effects independent of the disease process, or disease growth may be accelerated after progression. So, what is informative sensory? So, patients may be taken off study before progression for reasons such as toxicity, patient's or physician's preference, initiation of non-protocol therapy for various reasons. Informative censoring occurs when censored patients are at a different risk for treatment failure than those who remain on study. So, they are not censored at random. And on the right-side, we can see the picture of this. If the censor patients had higher risk for treatment failure, then, progression-free survival in this case is over-estimated. Another layer of complexity is added if informative censoring is differential or imbalanced, it occurs more often in the treatment than in the control group. And then, the efficacy of the intervention may be falsely over-estimated. A well reported and very educated example of informative censoring is BOLERO-2 clinical trial. BOLERO-2 clinical trial was a phase III randomised clinical trial, which randomised post-menopausal women to the group with everolimus and exemestane versus placebo and exemestane. And on the left-side of this slide, you can see data for progression-free survival. So, progression-free survival was improved by 6.5 months. So, hazard ratio was 0.36, and P was less than 0.001. Unfortunately, this difference didn't translate into the benefit in overall survival. Of note, 26 and 6.1 of the patients receiving everolimus and placebo respectively withdrew from the study prior to progression having mainly two treatment- related adverse events. Dr Templeton analysed censoring in trials in metastatic breast cancer. And he found that unbalanced censoring is quite prevalent. On the left-side of this slide, we see again, the data for progression-free survival as reported in BOLERO. And on the right-side of the slide, we see the data for time-to-treatment failure, analysed by Dr Templeton and colleagues and time-to-treatment failure counts discontinuation of treatment for any reason as an event. And we can see that this benefit of 6.5 months shrank to 1.1 month. Dr Gilboa et al. analysed informative censoring in oncology phase III clinical trials. In this analysis, 73 oncology randomised clinical trials were included and all these trials had significant surrogate time-to-event endpoint. 45% of these trials reported significant benefit in overall survival. These were so-called concordant trials. And 55% of trials did not report significant benefit in overall survival. These were discordant trials. Significant differential censoring in surrogates was 43% and 51% in discordant and concordant trials, respectively. Of note, significant censoring imbalance in the experimental arm occurred only in discordant trials, 15 versus 0%. While excessive sensory in control arm occurred more often in concordant trials, 28 versus 52%. After sensitivity analysis, 50% and 15% of the discordant and concordant trials lost their statistical significance in the surrogate. What are solutions to reduce bias of informative censoring? So, the primary endpoint should be overall survival rather than surrogate endpoint as informative censoring is not possible with overall survival. Analysing should be intention-to-treat. And we should use more often time-to-treatment failure as an endpoint, which could then be compared to other surrogate endpoints, such as progression-free survival. Reporting of censoring should be transparent. Under Kaplan-Meier curves. Nowadays, we see only data about the patients at risk under the Kaplan-Meier curves. And a possible solution is sensitivity analyses with worst- and best-case scenario. So,

what is association between progression-free survival and health-related quality of life in cancer patients? Hwang and doctor Gyawali did systematic view of 325 randomised clinic trials in advanced or metastatic solid tumours published between 2010 and 2015. And then, they found that only 54% of randomised clinical trials included quality of life endpoint and of these 147 randomised clinical trials reported quality of life outcomes. Association between progression-free survival and improvement in global quality of life was weak. When we are prescribing anti-cancer drugs, we are definitely dealing with harm. And this harm is unfortunately very often under-reported in clinical trials. So, what is the reason that harm is under-reported in clinical trials? Usually, the problems are symptomatic adverse events which are experienced by patients but under-reported by physicians, then investigators may detect some adverse events, but they are not reported properly, and short follow-up of the study may miss long-term toxicities and serious toxicities, which may only become apparent in post-marketing phase. But you are again encouraged to send some questions. So, how good are physicians in reporting of harm in clinical trials? Physicians reporting of symptomatic adverse events lacks reliability. Agreement between different physicians is moderate at best and different symptoms ratings can affect treatment decisions. Physicians under-report the incidence and severity of symptoms compared to reports of patients. Physicians' reports better predict unfavourable clinical events, such as death or emergency room visits than patients' reports. However, patients' reports better reflect underlying daily health status than clinical reports. Several years ago, we did a study in which we analysed updated drug labels of 12 targeted agents approved by the FDA. In these updated labels, we identified adverse drug reactions and then, we inquired initial drug labels and pivotal randomised clinical trials on the basis of which these drugs report. And as you can see on this slide, many adverse drug reactions or a substantial proportion of adverse drug reactions were not reported. So, this indicates that in the post-marketing period, it is very important to follow the possible harm, which anti-cancer drugs can cause. So, what is impact of the under-reporting of toxicity? So, patients do not know what symptoms to expect, and prolongation of life by potentially toxic treatments might not always be a priority for patients. For regulators and payers, they may not appropriately balance risks and benefits, and they may underestimate the cost of managing harm. Recently, an important step forward was done. So, instead of statistical significance, clinical value came to the forefront, and clinical value scale such as ESMO-MCBS scale emphasises improvements in duration and quality of survival. As we can see on this slide, ESMO scale has two scales. For curative settings, which includes grades A, B, C. A and B denotes the substantial benefit. And we have a scale for non-curative setting, grades from 1 to 5. Grades 4 and 5 denote substantial benefit. However, only a minority of studies evaluating drugs in non-curative setting demonstrate clinical value. In non-curative setting, only 38% of studies evaluating agents approved by EMA score 4 or 5. The story is different for curative setting where almost 95% of studies are evaluating agents approved by EMA score A. So, although these drugs are approved and we prescribe them to the patients, their clinical value may be trivial. Other relevant question is, do contemporary randomised control trials meet ESMO thresholds for meaningful clinical benefit? In this study, 277 randomised clinical trials evaluating systemic therapy for breast cancer, non-small cell lung cancer, colorectal cancer and pancreatic cancer, published between 2011 and 15, were reviewed. And experimental therapy was statistically superior to the control arm in 138 randomised clinical trials, which is a half of trials. Of these results of only 31% met the ESMO-MCBS clinical benefit threshold. Importantly, among the 226 randomised clinical trials for which the ESMO-MCBS could be applied, only 31% were designed to detect an effect size that could meet ESMO-MCBS thresholds. The last part of my talk I will dedicate to the efficacy-effectiveness gap. The efficacy-effectiveness gap addresses the difference in patient's outcome as reported in trials and in real world. In other words, we can ask ourself, can results from the pivotal clinical trial be applied to our patient in everyday clinical practise? Dr Templeton and colleagues did a very elegant study several years ago in which they compared median overall survival and toxicity of patients with advanced prostate cancer who were treated with three weekly docetaxel in routine clinical practise or in clinical trials at one institution, at The Princess Margaret Cancer Centre. And as we can see on this slide, median overall survival of patients who were treated in routine clinical practise was much less than for patients who were treated within clinical trials. In contrast, the toxicity in these patients who were treated in routine clinical practise was more

pronounced. This observation is pronounced by another study in which investigators compared more than 9.000 Medicare patients treated in everyday clinical practise with newly approved oncology drugs to more than 11.000 trial patients who were treated with experimental arm with same drugs within clinical trials. They found that median age in Medicare patients was much higher, which is not... this is expected as patients who are enrolled into Medicare programme should be 65-years old or order. They found that median overall survival among SEER-Medicare patients was shorter than among patients in the clinical trial intervention arm by 6.3 months. And on the left-side of this slide, we can see that this occurred in different cancers, like melanoma, like kidney cancer, breast cancer, and for different types of anti-cancer agents. Dose reduction or single prescription were more common among Medicare patients. Several years ago, we analysed eligibility criteria of randomised clinical trials. Into this analysis, we included 86 protocols written between 1987 and 2012. And we analysed these trials for eligibility criteria. We found that proportion of screened patients who were excluded from trials increased from 9 to 80% over a 10-year period. Increasing frequency of exclusion of patients were those who had prior cerebrovascular events, coagulation/bleeding disorders, prior gastrointestinal bleeding, cardio co-morbidities, and concurrent medication. So, decreasing frequency of exclusion in upper age limit usage and leukopenia. So, this study reinforces the fact that we should really worry about the efficacy-effectiveness gap in clinical trials. About our conclusions. For various reasons, our critical thinking can be compromised. So, if you are a resident, or if you're a young oncologist, I strongly encourage you to find a mentor in a highly academic institution and go for a fellowship at least for a while, for a few months. Cancer trials often do not directly measure and report patient-centred outcomes. While presentation of benefit of new anti-cancer therapy may be overly optimistic, harm may be under-reported in randomised clinical trials, and please, consider clinical value of new therapies and generalizability of results of randomised clinical trials in everyday clinical practise. Thank you very much for your attention.

Dr Templeton: Thank you so much Bostjan for this great talk. So, my name is Arnoud Templeton, and I'm happy to discuss questions. Please, do send them in using the A&Q function or the chat function. Maybe, I can start with a question, Bostjan. So, if you were the president of a wealthy pharma company, how would you design a clinical trial? Part-two will be, how would you design it if you were the president of a healthy academic centre?

Dr Seruga: So, as I mentioned, so, the selection of endpoint is definitely crucial, right. So, we should choose endpoints which are patient-centred and clear, this is overall survival and health-related quality of life. In the control arm, we should use treatment which is current standard. Sometimes in clinical trials, even in contemporary clinical trials, we see in the control arm treatments which are not quite standards, right. We have seen this recently in prostate cancer. But yes, I think that more often overall survival should be a primary endpoint in clinical trials, and health-related quality of life, definitely the secondary endpoint. And not only the secondary endpoint, but it should also be the results of this, endpoint should also be reported.

Dr Templeton: Yeah, so, I fully agree. I think what you mentioned is the crucial part for what really benefits our patients. So, they want to live better and/or longer. And I think that's something we should be able to show and otherwise, I'm not sure whether any treatment is really good for the patient. And yes, I also agree and I wonder what others in the session may think about this. I think looking at the control arm actually is indeed crucial because as you mentioned, we've seen quite a few studies recently, for example, in prostate cancer where the control arm pretty much was, well, maybe the weakest just still somehow acceptable version of treatment, not to say a toxic placebo. So, I think this is really crucial. And also, when we look at studies, when we read studies, I think asking ourselves, was this control arm really standard? Because if you compare to something very weak, the chance to see a difference is obviously bigger than when you compare to something that is stronger.

Dr Seruga: Yes. So, I would strongly encourage people to visit the ESMO-MCBS Scorecards, which are now freely available on the ESMO website to see what clinical value the drugs we prescribe in every day clinical practise actually have. We really may be surprised, because in a very busy clinic the decisions have to be...

you have to decide very rapidly, and maybe, we are over-prescribing some drugs, especially in countries where everything is available.

Dr Templeton: Yes. I fully agree. And actually, I think ESMO has been doing a very good job with incorporating these ratings in their guidelines. It is available. You just need to look at it. Now, still, Bostjan, let me ask you this. When you know, for example, for the everolimus/exemestane combination that, well, the risk-benefit ratio isn't beyond doubt. So, it doesn't get a very good score. Would you still use it in selected patients?

Dr Seruga: Well, honestly, I use it very rarely. So, the score now, as far as I know, is 2. If we take the results, if... of your re-analysis of time-to-treatment failure, where the benefit is 1.1 months. So, the initial score, I think, is 1 and if this is downgraded for toxicity, then the score is 0. Right?

Dr Templeton: Which probably comes quite close to something Professor Tannock once wrote in a critical letter. He said, well, if a drug adds toxicity and it doesn't prolong life, then, he considers this harmful. So, that's probably, well, putting it very strongly, but yeah, basically I agree. And I think, I really think we should encourage people and ourselves, first of all, to really consider whether there's a good risk-benefit ratio for that. These scoring systems are actually helpful.

Dr Seruga: You know, what I really miss in my practise, in the environment I work is some academic discussion about these things. Everybody is just talking about guidelines, what guidelines say. And every patient should have everything. And I think this is not a good service to our patients.

Dr Templeton: Do you think regulatory bodies like EMA or local health authorities should consider more clinical relevance, relevance scoring for clinical benefit rather than significant P-values for when approving a drug?

Dr Seruga: I think absolutely. You know, even these scores, ESMO-MCBS scores, which we have now, they don't take into account, for example, for PFS informative censoring, right? So, they take PFS as it is reported. So, this is... so, the scores for many of these new drugs are, I think, are over-inflated, over-estimated. And I definitely see, I definitely think we should be more restrictive in what we prescribe and what is of course approved.

Dr Templeton: Now, I think if a patient who comes in and says I want to really have this or that drug. And, you know, well, the scoring is 2 or 3. Would you discuss that with the patient? And if so, how would you discuss it with your patient?

Dr Seruga: I think it would be fair to discuss in a way we discuss now, right? So, this is not considered a substantial benefit. The benefit may be trial. The benefit is trivial in a trial. And when we bring this into the everyday clinical practise, the benefit can be even smaller. Because usually toxicity is greater and so on. So, I think we would need to discuss this with our patients, but, of course, for this we need more time in our clinics. And this is what I think many people are in a hurry and they have a lack of time and they cannot properly discuss these things with patients.

Dr Templeton: So, what can we do about this?

Dr Seruga: I think this is evolution of this, it's a good start that we have this now, clinical value scales, but I think these things will evolve over time. And in the future, I think we may get more restrictive in thinking what's approved and what's not approved in oncology.

Dr Templeton: I agree that lack of time to discuss things with a patient properly is a concern and is an issue. So, what might be at least one-step is when things are discussed internally in team. So, for example, we have weekly or biweekly journal club sessions where someone presents a paper, whatever it is, and then we have 10 minutes to discuss it afterwards. I think that that might be a starting point where questions can be brought up. Do you really think this progression-free survival is of any relevance to the patient? Do we know that the

patient benefit from this? Let's briefly have a look about at how the censoring ticks, and so on. I think that can be a starting point. And then, this might be a first step to sort of develop a first internal kind of mini consensus that it can be acceptable not to follow other guidelines for particular situations where there is, well, doubt about the objectivity or the clinical relevance really of the recommendation. What do you think?

Dr Seruga: I think absolutely agree with you. I think education is crucial. We should dedicate more time for education of younger generation, that they develop these skills. At least to try think critically about these things.

Dr Templeton: So, that's why we really hope many more people will watch your great lecture. Maybe, just one other question, maybe, you can just briefly repeat that. You said when you use overall survival as primary outcome, informative censoring cannot happen. Can you maybe just explain this a little bit more? I think it's very true, but please, comment a bit further on this. Why can't informative censoring happen?

Dr Seruga: Well, censoring means that the patient is lost to follow-up, right. But at the end, that occurs, right. And we get this information. Or did I understand your question well?

Dr Templeton: Yes. So, I think so, yes. So, maybe, I can try to say it in my words. So, you pointed out that one advantage of overall survival as primary outcome is that informative censoring cannot happen. And the reason basically is, as you just said, is that patients can only... they are only censored if they are lost to follow-up. So, what I like to use as an example when I discuss this, I say, well, a patient may go on holiday and fall in love and stay on in the nice place he went to and doesn't come back. So, he's lost for the study. So, the day he departed, at that point, he's censored. And this is very likely to happen at random in both groups equally. And then, censoring is not a problem. So, it's not censoring as such that is a problem, but it's a problem when it happens because of the intervention. That can be because of toxicity or when it's not equally balanced between the two arms. And because it's only because of loss to follow-up with survival, overall survival, as you said, is not happening. They're not a problem, they are censoring.

Dr Seruga: Yeah, we can find many examples in oncology trials, examples of informative censoring, unbalanced censoring which predominates in experimental arm.

Dr Templeton: Maybe one other question. When speaking about critical appraisal of the literature, when you come across a new article, how do you read the article?

Dr Seruga: Okay. So, the first thing I read is definitely abstract. So, how trial was designed, what was the primary endpoint? And then, of course, I check how these things are reported, right? And what are the conclusions? Because sometimes, and this is not so rare, we can find some spin in reporting of abstracts. So, every abstract should have reported the primary endpoint results. And then, we should also see some data about the toxicity of these new agents. And of course, then, I go into the body of the article and try to think these out in depth, right. Especially for disease sites, which I work. If this is something else then, I usually stay with abstract. I don't go into details in every published article.

Dr Templeton: So, the point you're making, I think it's is well taken. So, spin and bias in reporting results, especially in abstracts. I think that's truly a concern. Yes, so, I think that, for example, refers to when the primary outcome is negative, then, some secondary endpoint is highlighted for example. Is that what you mean?

Dr Seruga: Yeah. Yeah. So, many people are not in the academics, and they, I'm sure that as far as I know, for example, from my colleagues, they just check the abstract. And I think abstract is the most important part that's available for everybody, right? Because many people do not go into the details in the articles.

Dr Templeton: Yes, I agree. And especially, when we know that more and more papers are written by medical writers who are highly skilled professionals, but unfortunately paid by a pharmaceutical company. And we

all know that you don't want to bite the hand that feeds you. So, I guess that's part of the problem and increases the importance of being quite alert of how things are reported.

Dr Seruga: You know, very often I find useful information in supplementary materials, which is hidden. And not many people I think check supplementary materials online.

Dr Templeton: Yeah. And still, I think one of the good things really is more and more journals make supplementary material available, at least on their website. And also, the statistical analysis plan and the protocol has become more and more frequently readily available. So, I think that's quite good. Now, there are protocols that half of the protocol is just black, so you don't know what is written, but at least it's available. And I think that's certainly very good. But again, probably this is quite time consuming. And I can't say I read many protocols because yes, I just don't have the time, but at least it's available. And looking at how many in an exclusion criteria, or at least how many pages of in an exclusion criteria are in the protocol, I guess it's worthwhile because it gives you some measure of how representative for daily practise the study was.

Dr Seruga: You know, I really think that during the training, it's important that you write, that you do some research, even if this is observational research and you write the abstract. So, then, you realise what's important in the abstract, how to write the abstract. And you can use these skills then, of course, when you read reports of important pivotal clinical trials in oncology. So, I think it's really important that young people do some research to better understand how things are reported and done.

Dr Templeton: Yes. I couldn't agree more. Yeah. Okay. I think we are approaching 7:00 pm and we decided to close on time. I don't see any other questions. So, there's one minute left for further urgent questions. If they don't come up, I really would like to thank everyone who listens or looks at this session later on. And thank you so much for the great lecture, Bostjan, and have a wonderful evening and bye-bye to everyone. Thanks for looking. Bye. Bye-bye.

Dr Seruga: Thank you very much to you. Thank you. Bye-bye.