ESMO Recommendations on Homologous Recombination Deficiency Testing to Predict PARP Inhibitor Benefit in Ovarian Cancer

Supplementary Methods

Selection of working group members

Members of this working group comprise oncologists (REM, JAL, AL, IRC, EMS, LRY, DC, CS), a geneticist (SNZ), pathologist (XMG) and basic scientists (VS, SPS, DWG, CJL, JJ, DB). The consensus group was initiated by and includes members of the ESMO Translational Research and Precision Medicine Working Group (REM, LRY) and the ESMO Gynaecological Committee (JAL, IRC). WG members were selected based on expertise in two or more areas relevant to the topic.

Systematic review approach

PubMed searches (last accessed 24.09.2019) were performed to screen the literature for the range of HRD tests used in cancer research or clinical trials by identifying relevant reviews (search terms – 'PARP inhibitors' and 'HRD'), original research articles (search terms – ('HRD' OR 'BRCAness' OR 'homologous recombination') AND 'cancer' AND ('test' OR 'biomarker' OR 'assay') AND ('PARP inhibitor' OR 'cisplatin' OR 'carboplatin' OR 'oxaliplatin' OR 'olaparib' OR 'rucaparib' OR 'talazoparib')) and published clinical trials (individual searches for each PARPi (olaparib, rucaparib, talazoparib, niraparib and veliparib). The clinical trials database (https://clinicaltrials.gov) was interrogated (last accessed 16.01.2020) for additional relevant studies by selecting ovarian cancer ('condition or disease' search field) and HRD related terms ('HRD' OR 'homologous recombination') or individual PARP inhibitor names ('other terms' field) and filtering by completed phase II/III studies, with results. The citation lists of review articles were screened for additional relevant original research articles and publications related to commercially available tests were identified from company websites.

Using these strategies 343 relevant records were identified. For each record, as a minimum, the title and abstract was screened to identify the range of HRD tests in use/ development and to determine whether the record was appropriate for critical evidence appraisal of HRD test methodologies in ovarian cancer (see **Supplementary Table 1** for details). For critical evidence review records were identified that constituted original research in the context of ovarian cancer and included HRD biomarkers. The list of 52 potentially relevant studies was then reviewed and approved by all members of the expert panel who added 15 additional studies resulting in a total of 67 studies that were included in the evidence review. For each HRD biomarker test all shortlisted studies were categorised independently by at least two members of the WG using the level of evidence (LOE) approach and for genomics based tests using the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) ranking [18, 19].

The LOE approach as set out by Simon et al. was developed to provide a robust framework for assessing the quality of studies that include archival specimens to evaluate prognostic and predictive biomarkers [18]. After identifying relevant studies, we assigned each study to a category (A-D) according to criteria such as study design, specimen details, statistical design and validation. The overall LOE supporting a given test was then determined by taking into consideration consistency in results and the evidence category of each relevant study. Following the EGAPP system we assessed the quality of individual studies within the criteria of analytic validity, clinical validity and clinical utility. Studies were defined as good, fair or marginal where marginal reflects the fact that the study may not have been poor in general, but it may not have been designed to address our specific question [19].The main objective of the EGAPP approach is to determine whether there is a direct link between test use and a meaningful improvement in outcome or is useful in medical or personal decision-making. 'Direct evidence' is defined using the US Preventive Services Task Force (USPSTF) definition - a single body of evidence establishes the connection between the use of the genetic test (and possibly subsequent tests or interventions) and health outcomes [77].

The EGAPP model process for evaluating genomic tests defines clinical validity as the "accuracy of detection or prediction of phenotype, clinical disease or predisposition to disease". In the context of HRD tests, which do not necessarily directly test for a specific genetic variant, but rather are used to identify potential benefit from a treatment, which is itself an indirect measure HRD this definition needed to be adapted. Our agreed definition for assessing clinical validity of an HRD test is *"accuracy of prediction of PARP inhibitor benefit"*. The term benefit was preferred to sensitivity because many clinical trials were performed in the maintenance setting where we often are not able to measure objective disease response but we can measure PFS.

Supplementary Results and Discussion

The literature search identified a range of HRD biomarker tests in current clinical practice or development that fell in to 3 main categories: (i) HRR gene level tests (ii) Genomic signatures of HRD activity (including 'scars') and (iii) Real-time indicators or functional assays of HRD (**Figure 1**). Members of the expert panel performed critical review of the evidence relating to each of these categories. 19 studies were considered uninterpretable/ irrelevant to the exercise and excluded from further assessment. Ultimately a total of 47 informative studies were reviewed and assessed using the LOE and EGAPP approaches, where appropriate. The latter requires an assessment of analytic validity, clinical validity and clinical utility of each study. There were insufficient relevant clinical studies to apply a LOE to genomic scar assays, HRR-gene methylation and functional assays.

Additional Comments from the Panel Relating to HRD Testing Methods

HRR Genes

The analytic validity of *RAD51C*, *RAD51D* and other HRR mutations in terms of guiding the selection of therapy is somewhat stymied by the numbers of patients with mutations in these genes being included in clinical trials. As such the associations between, for example, PARPi sensitivity and these mutations is somewhat anecdotal but nevertheless consistent with all of the pre-clinical data that suggests patients with these mutations are likely to respond. However, it remains unclear whether all genes/effectors implicated in HRR are equal to HR competence. *BRCA1/2* loss leads to HRD and this is probably the case for *RAD51*, it is not clear for mutations and/ or deletions in other HR genes. Furthermore, individual mutations are not highly recurrent and therefore predicting the functional relevance of an individual point mutation or structural variant within a given gene footprint cannot usually be interpreted in isolation. Corroborating evidence of HRD from a genomic mutation/ scar test and/ or a functional assay would ideally be acquired.

Genomic scars / signatures

Conflicting data together with the fact that the test used in the relevant trials are commercial ones (with no re-imbursement in Europe) do not support their routine use for now. Ideally, continued work should be done to provide a more robust predictive biomarker, preferably a test developed and available on academic platforms. HRD scores should be more informative in the first line where we have less information on platinum sensitivity. Indeed, their positive predictive value, as measured indirectly by the HR versus placebo in the maintenance setting are favourable in PAOLA-1 and PRIMA. The NPV of these tests is difficult to interpret with PAOLA-1 showing no benefit in HRD-negative/unknown and PRIMA showing a benefit in HRD-negative [8, 13].

At present, none of the DNA sequencing approaches assess the presence of the known mechanism of clinical resistance, namely HR gene reversion.

Functional assays

Questions exist as to whether current functional assays can be applied clinically. The definition of thresholds that are highly predictive of benefit needs to be clarified and/or the likelihood that HR functionality of the whole tumour can be inferred from profiling of only one tumour section.

Key ethical, legal and social implications

If a validated test can be developed that accurately differentiates between HRD and HRP tumours and predicts PARPi benefit it would have important societal and cost implications by minimizing the number of patients exposed to costly and potentially toxic treatments. This is especially the case in the first line setting where the median PFS in placebo arm is over one year, which means potentially prolonged exposure to unnecessary medication for some. An ethical implication is the requirement to make a choice between a highly specific test, which will limit the number of patients eligible for PARPi treatment and thus be advantageous from a societal cost standpoint, versus a sensitive test, which will help to ensure patients who may benefit are not missed but will mean that we treat many unnecessarily. Incorporating patient group opinions in formulating these decisions will be important to determine

acceptance of exposure to daily tablets according to predicted levels of benefit. This is important when we consider that in breast cancer we may propose adjuvant strategies that improve survival by as little as 2-3%.

Comments from the panel relating to biomarker development

Some members of the panel suggested that existing commercial assays could be prohibitively expensive and data 'black boxes' that prevent further research. There was strong agreement, that provided a suitable assay for HRD testing existed, all patients with HGSOC should be considered for HRD testing in addition to *BRCA1/2* testing. All panel members felt this should be done at diagnosis and there was strong support for HRD testing at disease relapse, to guide PARPi therapy and when considering novel therapies / clinical trial entry.

WG discussions, questionnaire results and the systematic literature review identified major challenges that need to be addressed to develop more clinically useful tests for stratifying patients to PARPi use. These can be divided into the problems posed by cancer evolution and the need to acquire high-quality contemporaneous tumour samples and correlate with robust clinical data. The WG strongly supported the development of composite biomarkers.

Tackling Cancer Evolution

Cancers are patchworks of genetically related, yet distinct groups of cells termed subclones. Over time the subclone composition of a tumour changes driven by ongoing mutation and selection [78]. Through this process of cancer evolution, a HRP tumour can become HR deficient for example by acquiring a somatic mutation or methylation of a HR gene promoter region. Conversely a HRD tumour may become HRP by acquiring a reversion mutation in a HR gene [15, 16].

Current HRD tests do not provide a dynamic readout and are only valid for the time point at which the sample is obtained. In reality, the tested sample is usually archival, typically obtained at diagnosis or surgical debulking. If a HRD test is to be used to guide treatment at relapse or in the maintenance setting it ideally should be performed on a sample obtained at that point in time. Successful strategies would therefore need to tackle the associated problems of minimal residual disease and inter-tumour heterogeneity (at any point in time there may be multiple cancer subclones present). So called liquid biopsies, that sample circulating tumour cells or circulating tumour DNA may offer hope for addressing these problems. There is some promising data from prostate cancer that demonstrated the utility of ctDNA for identifying *BRCA1/2* mutations and reversion mutations and their correlation to PSA level [79] but clearly more nuanced HRD testing methodologies have not been explored in this way.

A second challenge is that all of the gene based and genomic assays, by definition, provide information on mutations acquired in the past. The footprints from mutational processes active early in tumorigenesis may not reflect contemporaneous activity of DNA repair mechanisms. Genomic scars will be detected within relapsed tumours even if they have developed treatment resistance. Importantly,

at present, none of the DNA sequencing approaches assess the presence of the known mechanism of clinical resistance, namely HR gene reversion. It seems logical that the known reversion events should be included in genomic assays and further research is needed to elucidate the full range of mechanisms of clinical resistance. Another route to tackling this problem might be measuring subclone specific mutational signatures, ideally in serial samples.

It remains that, based on the biology of the disease, we cannot rely on any of the existing tests to accurately predict whether HRD is extant at the time of treatment or not. It is highly likely that some form of functional assay will be required for this, probably in combination with other HRD-tests.

Acquiring high-quality tumour samples and data

One of the major challenges in HRD biomarker development is the collection of tumor tissue of adequate quality for analysis. Even within the first line maintenance PARPi trials using the commercially available Myriad MyChoice assays up to 18% of patients had an unknown HRD status due to either failed test or inconclusive /missing result [6, 8, 13]. Archival FFPE tissue is often used, which as discussed above is not always representative of the current tumour HRD status. Furthermore, formalin fixation and sub-optimal storage conditions can result in nucleic acid degradation, DNA crosslinking, base substitution artefacts and strand breaks [64]. Fresh frozen material is optimal for WGS and fresh tissue for functional assays. In this context, HRD biomarkers based on tissue biopsy techniques that enable serial and multiple spatially distinct samples or liquid biopsies are key. A collaborative effort is required by the academic community to generate and store high-quality whole genome data associated with robust clinical datasets. These should be widely available to allow testing of academically developed assays for ongoing research and biomarker validation.