

Master protocols in immuno-oncology: do novel drugs deserve novel designs?

Luca Mazzarella ^{1,2} Stefania Morganti,¹ Antonio Marra,¹ Dario Trapani,¹ Giulia Tini,² Piergiuseppe Pelicci,^{2,3} Giuseppe Curigliano^{1,3}

To cite: Mazzarella L, Morganti S, Marra A, *et al.* Master protocols in immuno-oncology: do novel drugs deserve novel designs? *Journal for ImmunoTherapy of Cancer* 2020;**8**:e000475. doi:10.1136/jitc-2019-000475

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jitc-2019-000475>).

Accepted 14 March 2020

ABSTRACT

The rapid rise to fame of immuno-oncology (IO) drugs has generated unprecedented interest in the industry, patients and doctors, and has had a major impact in the treatment of most cancers. An interesting aspect in the clinical development of many IO agents is the increasing reliance on nonconventional trial design, including the so-called 'master protocols' that incorporate various adaptive features and often heavily rely on biomarkers to select patient populations most likely to benefit. These novel designs promise to maximize the clinical benefit that can be reaped from clinical research, but are not without costs. Their acceptance as solid evidence basis for use outside of the research context requires profound cultural changes by multiple stakeholders, including regulatory bodies, decision-makers, statisticians, researchers, doctors and, most importantly, patients. Here we review characteristics of recent and ongoing trials testing IO drugs with unconventional design, and we highlight trends and critical aspects.

INTRODUCTION

The last decade has witnessed a rapid surge in the number of new immune-modulating drugs, which we termed next generation immune modulators (NGIMs), that have entered clinical development in immuno-oncology (IO).¹ These drugs encompass a large and variegated group of novel chemical structures including not only monoclonal antibodies, but also conjugated antibodies, antibody-like molecules, small molecules, interleukins and even microbial-mimetic compounds to activate innate immune system, for example via toll-like receptors (TLRs).¹ In this issue of the journal, many of these novel targets are reviewed in depth, including T-cell immunoglobulin and mucin-domain containing-3 (TIM-3), Nucleoside-binding with motif 1 (NKG2A), T cell immunoreceptor with Ig and ITIM domains, CD39, NKG2A, lymphocyte-activation gene 3 (LAG3) and MerTK. Preliminary results for many of these drugs are promising, as in the case of agents targeting LAG3, TIM-3, natural killers and others.²⁻⁴ NGIMs are usually employed in combination with classic checkpoint inhibitors, such as monoclonal antibodies directed against

cytotoxic T-lymphocyte antigen-4 (CTLA-4) or programmed of death-1 (PD-1) and its ligand (PD-L1), as they are often devoid of meaningful clinical activity when used alone.

Regulatory approval and commercialization of these agents require generally the demonstration of superior efficacy as measured through defined endpoints and with a statistical methodology that is deemed adequate for the clinical context. Traditionally, this was achieved by progressing through a three-phase route culminating with a randomized clinical trial that demonstrates some advantage (superior efficacy, inferior toxicity or cost) over what is considered standard of care at the time. However, some experts consider this canonical model of clinical experimentation less suited for developing novel immune modulators and especially their combinations.⁵⁻⁷ Reasons are manifold but we argue that they can be grouped in to three main categories: a perceived inadequacy of classical toxicity and efficacy endpoints, the increasing reliance on biomarkers, and the rapidity with which the treatment indications and disease classification are changing.

Here we will first review in further detail these three factors and how they influence trial design specifically in IO. We will then introduce the concept of master protocol and its defining features, highlighting how these novel designs can be better suited for NGIM development. We will finally provide examples of currently ongoing master protocols and briefly discuss *pro et contra* of existing trials.

Problems with classical endpoints

Toxicity

Immune-related adverse events (IrAEs) can develop with significant latency and have been systematically underestimated in trials, becoming evident only with late follow-up or pharmacovigilance studies, particularly for neurological, cardiologic and nephrological events.⁸⁻¹¹ Thus, the classical reliance on specific and predictable events (dose-limiting



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Division of Early Drug Development for Innovative Therapies, IEO European Institute of Oncology IRCCS, Milan, Italy

²Experimental Oncology, IEO European Institute of Oncology IRCCS, Milan, Italy

³Department of Oncology and Hematology, Università degli Studi di Milano, Milan, Italy

Correspondence to

Dr Luca Mazzarella;
luca.mazzarella@ieo.it

toxicity, DLT) within the first cycle is commonly inadequate to define the correct dose to be tested for efficacy. Still, it would be unfeasible to wait for such long time to determine the correct dose to be tested for efficacy. Interest is rising in immediately integrating safety and early antitumor activity (clinical responses, pharmacodynamic and pharmacokinetic parameters, markers of immune activation) in composite endpoints such as the optimal biologically active dose,¹² but this requires better understanding of drug activity in the preclinical phase, which will be facilitated by more sophisticated models of tumor–environment interaction such as organoids and humanized immunocompetent mice.¹³ Overall, as early DLTs and dose-dependent toxicities rarely occur with immunotherapy, the decision of the dose for expansion based merely on IrAEs is largely considered inappropriate.

Response criteria

IO agents typically can produce different patterns of response compared with standard systemic therapies,¹⁴ which cannot be properly evaluated by means of the conventional response evaluation criteria in solid tumors (RECIST V.1.1) criteria.¹⁵ The immune-RECIST (iRECIST) criteria have been developed in order to capture the unique pattern of response of the IO agents¹⁶ and take into account atypical patterns of responses such as pseudoprogressions and hyperprogressions.¹⁷ The distinction between ‘real’ progression and pseudo-progressions is an active field of research and renders even more important the identification of adequate and noninvasive biomarkers, especially those that are measured dynamically over time.^{18–21} However, it must be emphasized that pseudo-progression events are rare events, described in less than 5%–10% of patients receiving immune checkpoint inhibitors (ICIs).²² Notably, in clinical setting and trials, more than 50% of the patients receive treatment beyond progression, possibly underlining an overemphasis of the event and related expected benefit.²³ Overall, around 4% of patients with melanoma who were treated with ICI seem to derive a benefit in terms of disease response beyond progression, suggesting that a selected subgroup truly have an advantage. The coexistence of different systems for response evaluation poses additional problems for *interim* decisions (particularly relevant for master protocols, as discussed below). While these criteria are currently adopted in the majority of clinical trials investigating IO agents, their harmonization and standardization also for regulatory purposes and drug approval represent an absolute priority. The recent American Society of Clinical Oncology - Society for Immunotherapy of Cancer (ASCO-SITC) statement on trial reporting in immuno-oncology suggests to report responses according to standard RECIST criteria in parallel with the IO-specific response criteria like iRECIST in order to provide comparisons with prior clinical trials, ensuring consistency in the estimation of the benefit.⁶

Efficacy endpoints

Although some IO agents received approval on the basis of response rate in the early phases of clinical development,^{24–25} short-term efficacy endpoints like response rate and progression-free survival (PFS) are increasingly called into question. For instance, a meta-analysis on PD-1 inhibitors failed to demonstrate a correlation between PFS and Overall Survival (OS), suggesting an imperfect role of PFS as decisional parameter *ad interim*.²⁶ Moreover, an exploratory analysis from another work suggested a potential surrogacy of 6 month PFS rate for OS, while overall response rate (ORR) did not seem a robust surrogate endpoint.^{27–28} Additionally, reporting HRs might not properly represent the treatment benefit due to nonproportional hazards of survival and long-term survival with IO agents.^{29–30} Other statistical models, including restricted mean survival time analyses and Cox regression models with arm–time interaction coefficients, have been proposed as more appropriate for evaluating survival outcomes in IO trials.³⁰ More recently, a Food and Drug (FDA) analysis of 10 randomized controlled enrolling patients with melanoma treated with ICI, assessed the relation between the depth of response and survival.³¹ At 24 months, 92% of the patients reaching a maximal tumor shrinkage of 75% or more were alive, suggesting a role of profound responders in the identification of patients deriving a greatest benefit. Finally, patient-reported outcomes are likely to gain increasing importance given the peculiar trade-off between toxicity and efficacy of IO agents, but still suffer from poor standardization as a recent review of IO studies reported.³²

Biomarkers and the trade-off between sample size and specificity

Experience with targeted agents has shown that early implementation of biomarker-based stratification can enhance the clinical benefit and the likelihood of successful drug approval.^{33–35} Translational studies with first-generation ICIs against PD-(L)1/CTLA-4 have revealed that also in the IO context, biological parameters can strongly influence efficacy or predict toxicity.³⁶ These factors include mutational signatures and specific mutations obtainable from DNA sequencing of the tumor sample or circulating free DNA,^{37–42} transcriptional signatures,^{43–45} circulating factors (neutrophil to lymphocytes ratio, LDH and others^{46–47}), cytokines,⁴⁸ intratumoral expression of immune markers (PD-L1, infiltrating lymphocytes, myeloid cell infiltration and others^{49–50}), early changes in peripheral immune cells²¹ or more sophisticated immune parameters like T-cell receptor repertoire changes,^{51–52} radiomic patterns,^{18–20} and the gut microbiota.^{53–55}

Enrichment for biomarker-positive subgroups is a logical approach to achieve large effect size in clinical trials. The overall viability of the chosen biomarker can inform trial design in key ways. If a biomarker is easily implemented in the workflow and/or has strong evidence for a good negative or positive predictive values and/or has sizeable prevalence, it may be advisable to implement

it early and solidly in the design, possibly in the inclusion criteria, to immediately define the treatment space for the investigational drug and limit costs. If these conditions do not hold, and especially if biomarker negativity would not fully justify exclusion from treatment, it may be advisable to only include it as an exploratory endpoint or use for stratification. The decision to restrict clinical development to biomarker-selected populations may be key to delimiting the clinical space in which that drug will operate, as very well exemplified by the different results (and consequent FDA approvals) of nivolumab (tested on patients unselected for PD-L1 expression) and pembrolizumab (tested on patients expressing PD-L1 on $\geq 1\%$ cancer cells) in first-line non-small cell lung cancer (NSCLC).^{56 57} Three main factors affect the clinical viability of a biomarker: performance (positive/negative predictive value and sensitivity), feasibility (technical, economical, ethical) and prevalence.

Performance

For many biomarkers currently tested in IO trials, predictive value is still speculative. More complex is the discussion over biomarkers for NGIMs. Some of these drugs seem to perform better when their molecular target is highly expressed (eg, LAG3) but given their indirect mechanism of action on the tumor, it is not surprising that local target expression correlates sub-optimally with efficacy. Probably, a combinatorial approach with probabilistic algorithms integrating multiple data dimensions will be required,⁵⁸ but this of course will require a trade-off with feasibility, as multiple data will require multiple laboratory tests. Still poorly defined are early markers of immune toxicity, an issue of great importance given the peculiar toxicity profile of these drugs⁴⁸

Feasibility

Biomarker feasibility can be affected by several factors. Technical complexity varies significantly from simple but poorly predictive tests to complex transcriptional signatures or dynamically assessed. Importantly, biomarkers require biological specimens, which can not always be collected, especially in specific cases like dynamic biomarkers requiring serial sampling. Sampling issues may be at least partially overcome by the growing capabilities of liquid biopsies, which have for instance shown to be an adequate surrogate of conventional biopsies for some biomarkers.^{39 59}

Prevalence

Targeted therapies offer numerous examples of drugs with exceptional efficacy in exceptionally rare populations (eg, NTRK-mutated or RET-mutated). Given the relatively low rate of long-term responders in IO-treated patients,^{60 61} it is not unlikely that responses could be predicted by single or combined biomarkers, each with extremely low prevalence. This may be of particular importance in multi-arm trials where unequal biomarker co-occurrence can lead to unbalanced arms.⁶²

Clinical research in a rapidly changing landscape

The advent of targeted and immune therapies has caused rapid and dramatic changes in the criteria for disease classification and treatment. The gradual shift towards a histology-agnostic, biomarker-driven disease classification, best exemplified by the approval of anti-PD1 agents for mismatch repair-deficient cancers irrespective of the tissue of origin,⁶³ has created novel clinical entities in which preliminary data to formulate testable hypotheses (eg, historical controls) are not or are poorly available. This can have significant implications on the design of clinical trials, when the pre-test assumptions are established. Also, the increasing adoption of IO in the early settings^{64–66} creates a growing cohort of patients with a history of treatment with ICI. How prior failure to ICIs can influence subsequent response to NGIMs is an open question. On one hand, subsequent success rate may be predicted to be lower if the patient has already progressed to an initial IO line. On the other hand, prior failure may itself be due to the induced upregulation of additional checkpoints that may be specifically targeted in the second line. Besides, IO sequencing may be relevant for toxicity. Unfortunately, preclinical models are mostly inadequate to study these issues, as IO sequencing is difficult to model.

Master protocols as an overarching framework for testing multiple hypotheses

The factors described above critically undermine the canonical model in specific ways. In phase 1, where the aim of the investigation is to establish safety and a recommended phase 2 dose, the opportunity to increase dosage is classically evaluated based on DLTs identified within the first cycle. If however toxicities are manifested with a delayed kinetics and with little dose-dependency, the discriminatory ability of classical '3+3' designs is severely impaired and may lead to significant underestimation of toxicity rates and difficulty in identifying the optimal dose.⁶⁰ In phase 2, when preliminary efficacy should be evaluated, three factors are of major importance: (1) the lack of a direct relationship between short-term (response rate and PFS) and long-term (overall survival) endpoints; (2) the increasing difficulty in identifying an appropriate historical control to gauge efficacy, given the rapidly changing treatment landscape; and (3) the need for biomarkers, which complicates screening procedures and further restricts inclusion criteria, making it more difficult to subsequently extend the results to wider patient populations. Finally, randomized phase 2 and 3 trials become hard to carry out because (1) biomarker stratification restricts the fraction of suitable patients, increasing dramatically the number of patients needed to screen; and (2) the choice of an adequate control arm becomes difficult and sometimes ethically questionable, as the standard of care may be rapidly changing and/or radically different in terms of expected outcome and quality of life.

Master protocols have arisen as a way to overcome these issues and provide statistical and organizational tools to test multiple hypotheses efficiently. Master protocols encompass trials designed to evaluate single drugs across multiple populations ('basket trials'), multiple drugs on a single population ('umbrella trials'), or complex multi-arm, multi-stage designs that evaluate multiple treatments simultaneously, also referred to as 'platform trials'.⁶⁷ In a certain sense, basket and umbrella trials can be considered special cases of master protocols/platform trials where number of tested drugs or populations is equal to 1. Populations can be defined by classical anatomical/histological parameters or by more sophisticated biomarkers. Examples of trials with multiple biomarker-defined populations include the I-SPY2 trial in the neoadjuvant setting for breast cancer, which includes gene expression signatures among the parameters used to stratify and adapt randomization,⁶⁸ or the Molecular Analysis for Therapy Choice of the National Cancer Institute (NCI-MATCH trial),⁶⁹ that uses a histology agnostic approach based on actionable mutations.

The use of a single infrastructure to evaluate multiple drugs and populations has the potential for efficient and accelerated drug development, reducing the exposure of patients to less effective or unacceptably toxic interventions.⁷⁰ In trials with multiple biomarker-defined arms, simultaneously testing for all biomarkers within the same framework minimizes the number of patients needed to screen and thus maximizes the chances of each patient to receive treatment within the trial.⁶² Furthermore, if the same drug or combination is used across multiple biomarker-determined groups with staggered enrolment, toxicity events can still be cumulated across arms, obtaining a more faithful description of the safety profile. On the other hand, the statistical design becomes extremely complex and relies critically on interim analyses when efficacy/futility within and across treatment arms is determined; this intermediate evidence is used to decide whether to continue, stop early or add new treatment options. Statistical strategies to modify *in medias res* the design of a clinical trial are collectively called 'adaptive designs'. For the FDA, the defining property of an adaptive design is the possibility to '*adjust to information that was not available when the trial began*'. In some cases, new evidence from other trials or from the same trial can be used to adjust the design of the platform, in opposition to the traditional trial design, where all the parameters are prespecified.⁷¹ The type of information that can be used to guide trial conduct can in principle be of heterogeneous nature and include efficacy, toxicity or a composite of the two, in the overall population or in biomarker-defined groups. Specific designs have been devised to continuously reassess efficacy and/or toxicity and adapt recruitment accordingly; these continuous reassessment designs may be better suited to adequately capture efficacy and toxicity of IO agents, with their peculiar, delayed kinetics.^{12 72–78} Master protocols do not strictly require adaptive designs, but in general a certain

degree of design flexibility is desirable to accommodate the heterogeneity associated with multiple patient populations. A peculiar feature of platform trials is the possibility for patients to be each counted multiple times within the same trial, since they are assessed for eligibility once and can potentially remain in treatment within multiple arms of the same trial. However, paradoxically this may limit the possibility for some therapies to be proven effective, especially if the multiple agents tested have common mechanisms of action and, consequently, common mechanisms of resistance, a plausible scenario for IO combinations.

Regulatory bodies are increasingly aware of the shifting paradigms in clinical research but also of the need to provide clear frameworks for the road to drug approval. Thus, FDA launched initiatives to help streamline alternative regulatory pathways, like the breakthrough designation pathway. These novel regulatory pathways have been exploited frequently in recent years, representing 43% of the novel FDA approvals between 2012 and 2017.⁷⁹ However, these expedite tracks are mostly still based on conventional trial design, with either short-term endpoints or nonrandomized but sufficiently large trials. Recognizing the lack of regulatory approvals based on complex innovative trial designs (CIDs), the FDA recently launched the CID pilot project.⁸⁰ CIDs are defined loosely to include multiple novel design techniques incorporating adaptive and Bayesian statistics, innovative use of external or historical control data, formal incorporation of prior knowledge into the study design (in this context, biomarkers and/or treatment history) and adaptive designs allowing prespecified modifications to the designs as evidence accumulates.⁸⁰

Some forms of adaptive trials (eg, Simon's two-stage or its derivations⁸¹) have entered widespread application, but are not yet considered sufficient evidence for regulatory approval. Many authors have raised concerns over the actual utility or even ethical soundness of adaptive designs.^{82 83} Most obviously, unplanned variations in sample size and inclusion criteria affect the possibility to interpret, validate and generalize results. Furthermore, although part of the rationale for adaptive designs is to reduce the number of patients on trial, the lack of fixed sample sizes may actually result in larger, rather than smaller sample sizes.^{82 83}

Mapping of recent or ongoing master protocols in IO

To map the current trends in trial designs for NGIMs, we reviewed clinical trials using a loose definition of an IO-centered master protocol combining different drugs. We included phase 1 and 2 interventional clinical trials in our analysis in which:

1. Three or more drugs were tested.
2. Two or more combinations of different agents were considered, including at least one IO agent.
3. At least two different immunotherapy drugs were included, of which at least one was ICI.

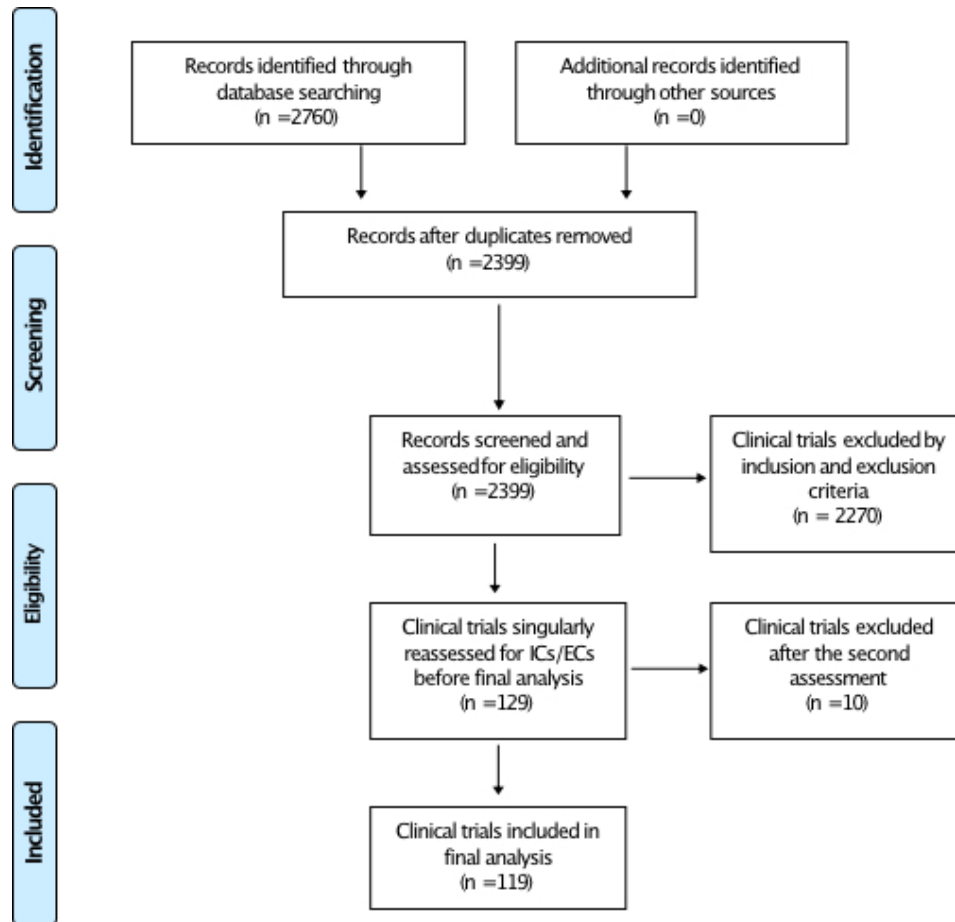


Figure 1 PRISMA diagram describing criteria chosen for literature review. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses

We excluded trials that provided for one or more arms in which chemotherapy was administered alone, with the exception of studies considering more than two arms. This choice was made to exclude all trials that compared chemotherapy to a unique combination of the same chemotherapy plus an IO agent. Trials including adoptive therapies and vaccines were not considered. We excluded also trials that enrolled only hematological cancers.

This results in 119 trials being considered for review (figure 1). Most are phase 1 (n=51) or phase 1/2 (n=41) and are nonrandomized. In online supplementary table 1, we categorized these trials based on setting (mostly advanced and metastatic), treatment history (prior IO allowed, required or not allowed); the use of biomarkers (prospectively included), and the timing of their assessment (baseline vs pretreatment or posttreatment), and scope (for inclusion, stratification or exploratory); the presence of adaptive features; if the combination strategy included IO drugs only or also other modalities like chemotherapy, targeted agents, or radiation therapy. Specific methodology is included in the online supplementary file 1.

The analysis shows a dramatic rise in the total number of trials and patients enrolled in IO-centered master protocols (figure 2A), the increasing reliance on biomarkers

(figure 2B) and the allowance or even requirement for prior IO from 2016 (figure 2C). Most likely these trends were pushed by the approval of anti-PD1 agents nivolumab and pembrolizumab in melanoma and NSCLC, occurred in late 2014/early 2015. These agents represent the most common backbone for combination therapies, together with ipilimumab and durvalumab (figure 3A).

The number of agents being tested in each trial is most commonly three, but some trials test even more than 10 drugs at the same time (figure 3B).

Statistical design of the trials

Adaptive features in the design of most trials are difficult to evaluate as complete statistical design is not routinely reported in public databases, a practice that will probably change in light of the more positive attitude of regulatory bodies toward adaptive designs, demanding more transparency in the conduct of clinical trials and disclosure of data (figure 3C). In our study, 47% (56/119) of the studies did not publicly disclose sufficient elements of their statistical design, 23% (28/119) use nonadaptive designs, 12% (14/119) included some adaptive features that were not disclosed, and only 18% (21/119) reported verifiable adaptive features (mostly Simon's two-stage design). Unsurprisingly, long-term outcomes like overall

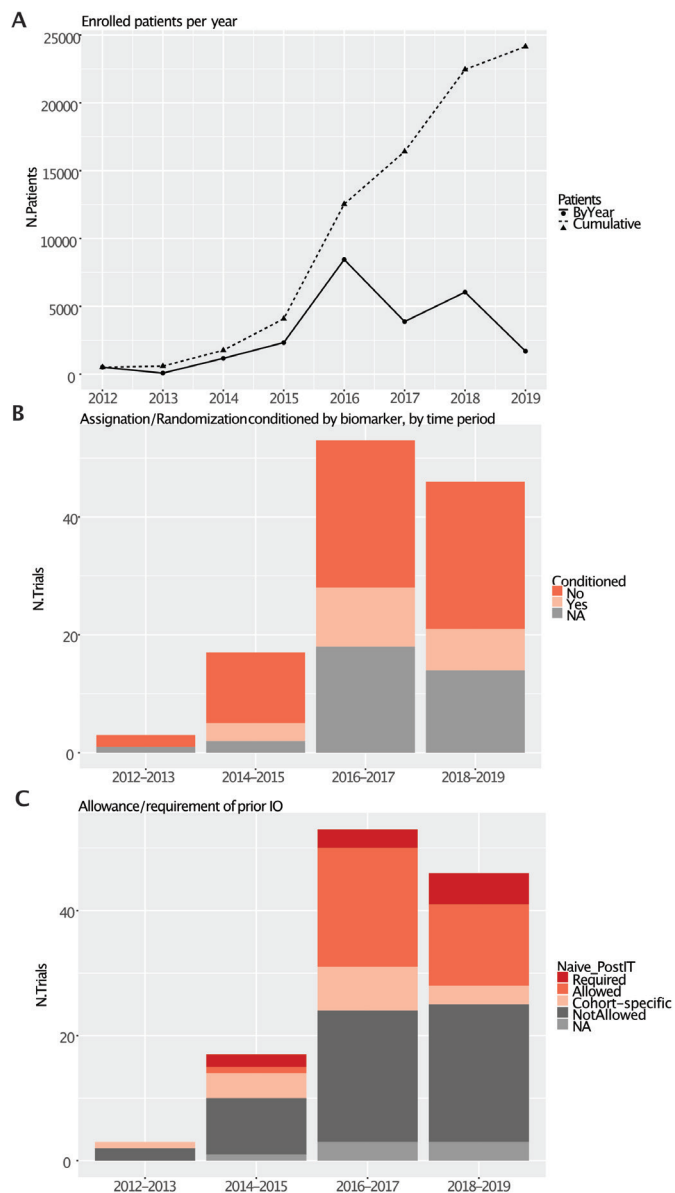


Figure 2 General characteristics of papers included in the review. (A) Yearly and cumulative planned patient enrolment over time. (B) Trial classification by biomarker use by biennia (C) Allowance/requirement of prior IO by biennia.

survival very rarely constitute the primary endpoint (figure 3D), perhaps related to the prominence of earlier phase of trials.

Specific examples

The most paradigmatic examples of ongoing IO master protocols are those centered around single pharmaceutical companies' and their compounds, mainly based on anti-PD-L1 drugs: BMS' FRACTION (nivolumab)⁸⁴ Roche's MORPHEUS (atezolizumab)⁸⁵ and Merck/Pfizer's Javelin Medley (avelumab).⁸⁶

FRACTION

FRACTION is the only platform whose design is published and described in detail. FRACTION incorporates all elements discussed above: multiple NGIMs (including

anti-CTLA4, LAG3, IDO1, cytokines and others) are tested in combination with a nivolumab backbone, in multiple patient populations defined by diseases (lung, gastric and renal cell cancer), biomarkers and prior treatment history. Several adaptive features are incorporated, including the possibility for internal crossover and premature arm termination or expansion. Interim decisions are not solely based on short-term efficacy but also include safety and pharmacodynamics. Biomarkers are investigated both as inclusion criteria (PD-L1) and as exploratory endpoints (gene expression signatures, mutational burden, immune infiltration, oncogenic mutations, germline makeup and peripheral markers). An interesting feature of FRACTION is its 'modular' architecture: alongside a 'backbone' master protocol defining general rationale and common inclusion criteria, study conduct for each specific disease is regulated by subprotocols, an agile structure that allows the introduction of modifications by amending the subprotocol without perturbing the master protocol.

MORPHEUS

MORPHEUS is a platform with disease-oriented multiple arms of study. Multiple combinations of atezolizumab and other IO and non-IO agents or chemotherapy are assessed for NSCLC, triple-negative breast cancer, colorectal, gastric and pancreatic cancer and transitional-cell malignancies of the urinary tract. The published primary endpoint of the arms to assess the activity of the combination regimens is ORR, across the studies. Secondary endpoints explore longer-term outcomes, including PFS and OS.^{85 87}

Javelin Medley

Javelin Medley was designed as an open-label, phase 1b/2 study evaluating the combination of avelumab with multiple combinations of IO and other agents. The goal of this study was to determine safety, tolerability, and clinical activity of these combination therapies, based on endpoints of safety (DLT) and activity (ORR).⁸⁶ Avelumab is combined with multiple agents across several immunecheckpoints anti-4 IBB (utomilumab), M-CSF (PD 0360324), agonist of OX-40 (PF-04518600) and TLR9 (CMP-001), in doublets or triplets.

Advantages and disadvantages of master protocols in IO

Advantages of complex, multilayered designs like FRACTION's are many. The amount of information that can be squeezed out from a single study increases dramatically, including longitudinal information after crossover that is not typically collected in traditional trials and that may inform subsequent strategies for treatment sequencing. The administrative work is minimized compared with what would be needed if each hypothesis was tested independently. Perhaps most importantly, likelihood of exposing each patient to the best therapy is increased, both because clearly ineffective or toxic arms are promptly terminated and because crossover allows to

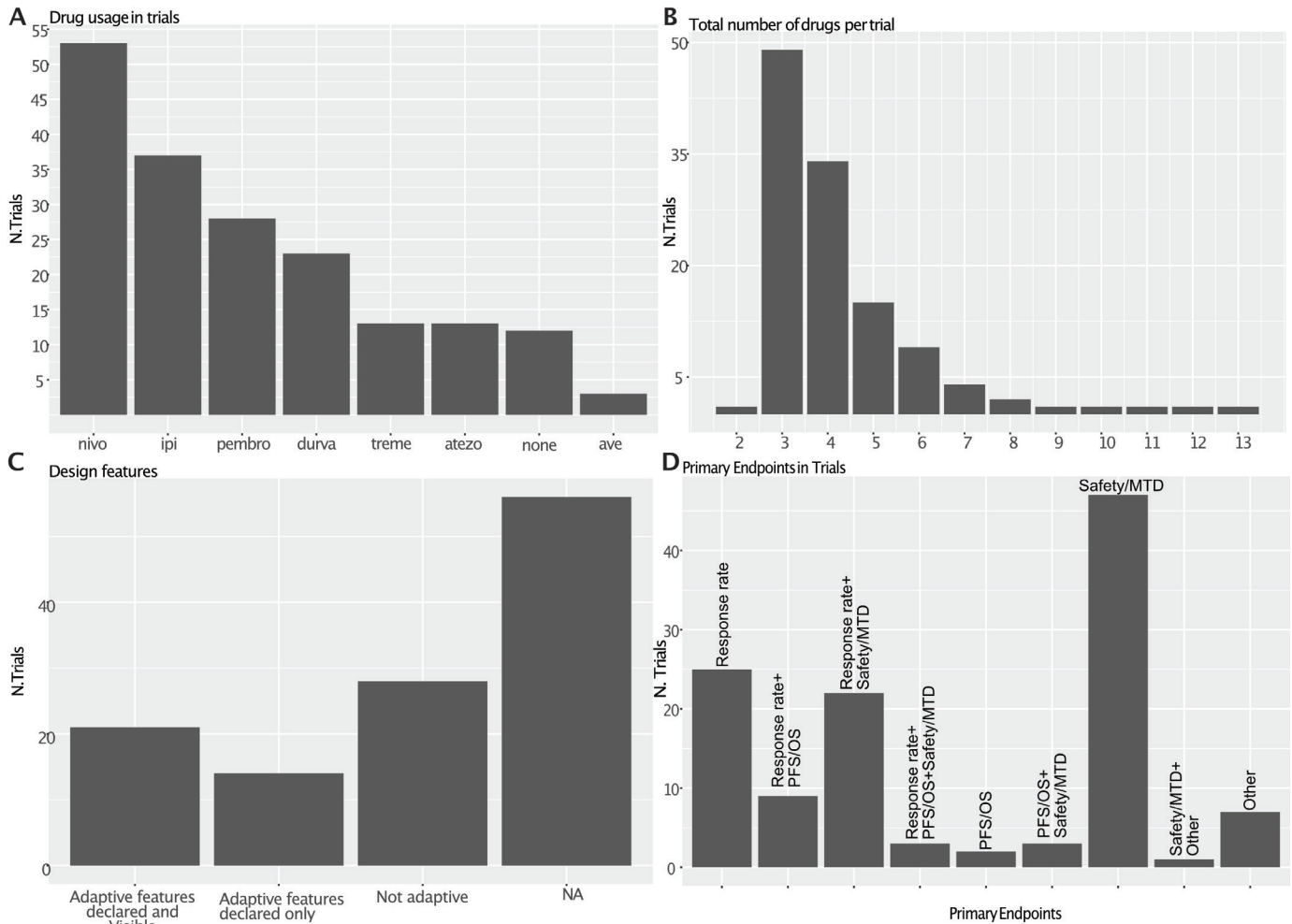


Figure 3 Specific characteristics of papers included in the review. (A) Drug inclusion. (B) Distribution of total number of drugs per trial. (C) Inclusion of adaptive features in the design. (D) Primary endpoints: is it a crossover when the patient moves in another arm?. MTD = Maximum Tolerated Dose

receive multiple treatments that also take into account initial responses.

However, some important prices are to be paid. Statistical strength is sacrificed, because of small arm size, frequent crossover, and poor reproducibility as decision criteria are variegated; thus, results are always highly exploratory, and the risk of an inconclusive or misleading trial is high,^{83 88} as experienced in the field of targeted therapy where such novel designs have been explored first. For instance, I-SPY 2 predicted 88% probability of success in a phase 3 trial for neoadjuvant carboplatin and veliparib, but this was not confirmed in the randomized controlled phase 3 clinical trial BrighTNess.⁸⁹

CONCLUSIONS

As we have shown above, reliance on nonconventional, complex trial designs, especially at the intermediate stages of drug development, is increasing noticeably. Such designs have been called into question from a methodological and even ethical standpoint.^{83 88} An argument can be made, however, that judging the success of a trial solely on the basis of a primary survival endpoint can

be reductive, especially for disease eligible to multiple lines of therapies as platinum-sensitive ovarian cancer or endocrine-positive breast cancer, in a landscape with ever-multiplying treatment options and a pressing need of criteria for adequate patient selection. Finally, increasing patient awareness raises pressure from the final users to gain access to therapies. Time will tell if the recent opening to nonconventional designs by FDA will smoothen or ignite the well-known ‘access vs evidence’ regulator's dilemma.⁸²

Twitter Luca Mazzarella @matsadoc

Contributors LM and GC conceived the study. LM, SM, AM, DT, GT, PP and GC collected and analyzed data. LM and GT performed statistical analysis. All authors contributed to writing. All authors read and approved the final manuscript.

Funding Research in LM's lab is funded by the EU ERAPerMed JTC2018 - PEVOSQ-data project.

Competing interests GC received honoraria for speaker, consultancy or advisory role from Roche, Pfizer, Novartis, Seattle Genetics, Lilly, Ellipses Pharma, Foundation Medicine, and Samsung.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Luca Mazzarella <http://orcid.org/0000-0002-4043-3279>

REFERENCES

- Mazzarella L, Duso BA, Trapani D, *et al*. The evolving landscape of 'next-generation' immune checkpoint inhibitors: A review. *Eur J Cancer* 2019;117:14–31.
- Diamond JR, Standifer N, Ascierto ML, *et al*. Translational endpoints in patients with metastatic microsatellite-stable colorectal cancer (MSS-CRC) treated with durvalumab plus monalizumab (anti-NKG2A). *Ann Oncol* 2018;29(Suppl 8):viii425.
- Curigliano G, Gelderblom H, Mach N, *et al*. Abstract CT183: phase (Ph) I/II study of MBG453± spartalizumab (PDR001) in patients (pts) with advanced malignancies. *Cancer Res* 2019;79:CT183.
- Ascierto PA, Bono P, Bhatia S, *et al*. LBA18Efficacy of BMS-986016, a monoclonal antibody that targets lymphocyte activation gene-3 (LAG-3), in combination with nivolumab in pts with melanoma who progressed during prior anti-PD-1/PD-L1 therapy (mel prior IO) in all-comer and biomarker-enriched populations. *Ann Oncol* 2017;28:mdx440.011.
- Adashek JJ, LoRusso PM, Hong DS, *et al*. Phase I trials as valid therapeutic options for patients with cancer. *Nat Rev Clin Oncol* 2019;16:773–8.
- Tsimberidou AM, Levit LA, Schilsky RL, *et al*. Trial reporting in immuno-oncology (TRIO): an American Society of Clinical Oncology-Society for immunotherapy of cancer statement. *J Clin Oncol* 2019;37:72–80.
- Postel-Vinay S, Aspeslagh S, Lanoy E, *et al*. Challenges of phase 1 clinical trials evaluating immune checkpoint-targeted antibodies. *Ann Oncol* 2016;27:214–24.
- Weber JS, Hodi FS, Wolchok JD, *et al*. Safety profile of nivolumab monotherapy: a pooled analysis of patients with advanced melanoma. *J Clin Oncol* 2017;35:785–92.
- Salem J-E, Manouchehri A, Moey M, *et al*. Cardiovascular toxicities associated with immune checkpoint inhibitors: an observational, retrospective, pharmacovigilance study. *Lancet Oncol* 2018;19:1579–89.
- Johnson DB, Manouchehri A, Haugh AM, *et al*. Neurologic toxicity associated with immune checkpoint inhibitors: a pharmacovigilance study. *J Immunother Cancer* 2019;7:134.
- Ali AK, Watson DE. Pharmacovigilance assessment of immune-mediated reactions reported for checkpoint inhibitor cancer immunotherapies. *Pharmacotherapy* 2017;37:1383–90.
- Wages NA, Chiuhan C, Panageas KS. Design considerations for early-phase clinical trials of immune-oncology agents. *J Immunother Cancer* 2018;6:81.
- Buqué A, Galluzzi L. Modeling tumor immunology and immunotherapy in mice. *Trends Cancer* 2018;4:599–601.
- Wolchok JD, Hoos A, O'Day S, *et al*. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res* 2009;15:7412–20.
- Tazdait M, Mezquita L, Lahmar J, *et al*. Patterns of responses in metastatic NSCLC during PD-1 or PDL-1 inhibitor therapy: comparison of RECIST 1.1, irRECIST and iRECIST criteria. *Eur J Cancer* 2018;88:38–47.
- Seymour L, Bogaerts J, Perrone A, *et al*. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol* 2017;18:e143–52.
- Borcoman E, Nandikolla A, Long G, *et al*. Patterns of response and progression to immunotherapy. *Am Soc Clin Oncol Educ Book* 2018;38:169–78.
- Sun R, Limkin EJ, Vakalopoulou M, *et al*. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* 2018;19:1180–91.
- Niemeijer AN, Leung D, Huisman MC, *et al*. Whole body PD-1 and PD-L1 positron emission tomography in patients with non-small-cell lung cancer. *Nat Commun* 2018;9:4664.
- Bensch F, van der Veen EL, Lub-de Hooge MN, *et al*. (89)Zr-atezolizumab imaging as a non-invasive approach to assess clinical response to PD-L1 blockade in cancer. *Nat Med* 2018;24:1852–8.
- Huang AC, Postow MA, Orlowski RJ, *et al*. T-Cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature* 2017;545:60–5.
- Chiou VL, Burotto M, Pseudoprogession BM. Pseudoprogession and immune-related response in solid tumors. *J Clin Oncol* 2015;33:3541–3.
- Beaver JA, Hazarika M, Mulkey F, *et al*. Patients with melanoma treated with an anti-PD-1 antibody beyond RECIST progression: a US food and drug administration pooled analysis. *Lancet Oncol* 2018;19:229–39.
- Garon EB, Rizvi NA, Hui R, *et al*. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med* 2015;372:2018–28.
- Chow LQM, Haddad R, Gupta S, *et al*. Antitumor activity of pembrolizumab in Biomarker-Unselected patients with recurrent and/or metastatic head and neck squamous cell carcinoma: results from the phase Ib KEYNOTE-012 expansion cohort. *J Clin Oncol* 2016;34:3838–45.
- Gyawali B, Hey SP, Kesselheim AS. A comparison of response patterns for progression-free survival and overall survival following treatment for cancer with PD-1 inhibitors: a meta-analysis of correlation and differences in effect sizes. *JAMA Netw Open* 2018;1:e180416.
- Ritchie G, Gasper H, Man J, *et al*. Defining the most appropriate primary end point in phase 2 trials of immune checkpoint inhibitors for advanced solid cancers: a systematic review and meta-analysis. *JAMA Oncol* 2018;4:522–8.
- Hamada T, Kosumi K, Nakai Y, *et al*. Surrogate study endpoints in the era of cancer immunotherapy. *Ann Transl Med* 2018;6:S27.
- Anagnostou V, Yarchoan M, Hansen AR, *et al*. Immuno-oncology trial endpoints: capturing clinically meaningful activity. *Clin Cancer Res* 2017;23:4959–69.
- Alexander BM, Schoenfeld JD, Trippa L. Hazards of hazard ratios—deviations from model assumptions in immunotherapy. *N Engl J Med* 2018;378:1158–9.
- Osgood C, Mulkey F, Mishra-Kalyani PS, *et al*. FDA analysis of depth of response (DpR) and survival across 10 randomized controlled trials in patients with previously untreated unresectable or metastatic melanoma (UMM) by therapy type. *J Clin Oncol* 2019;37:9508.
- King-Kallimanis BL, Howie LJ, Roydhouse JK, *et al*. Patient reported outcomes in anti-PD-1/PD-L1 inhibitor immunotherapy registration trials: FDA analysis of data submitted and future directions. *Clin Trials* 2019;16:322–6.
- Schwaederle M, Zhao M, Lee JJ, *et al*. Association of biomarker-based treatment strategies with response rates and progression-free survival in refractory malignant neoplasms: a meta-analysis. *JAMA Oncol* 2016;2:1452–9.
- Schwaederle M, Zhao M, Lee JJ, *et al*. Impact of precision medicine in diverse cancers: a meta-analysis of phase II clinical trials. *J Clin Oncol* 2015;33:3817–25.
- Jardim DL, Fontes Jardim DL, Schwaederle M, *et al*. Impact of a biomarker-based strategy on oncology drug development: a meta-analysis of clinical trials leading to FDA approval. *J Natl Cancer Inst* 2015;107. doi:10.1093/jnci/djv253. [Epub ahead of print: 15 Sep 2015].
- Zaretsky JM, Garcia-Diaz A, Shin DS, *et al*. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *N Engl J Med* 2016;375:819–29.
- Shin DS, Zaretsky JM, Escuin-Ordinas H, *et al*. Primary resistance to PD-1 blockade mediated by JAK1/2 mutations. *Cancer Discov* 2017;7:188–201.
- Conway JR, Kofman E, Mo SS, *et al*. Genomics of response to immune checkpoint therapies for cancer: implications for precision medicine. *Genome Med* 2018;10:93.
- Gandara DR, Paul SM, Kowanetz M, *et al*. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med* 2018;24:1441–8.
- Coelho MA, de Carné Trécesson S, Rana S, *et al*. Oncogenic RAS signaling promotes tumor immunoresistance by stabilizing PD-L1 mRNA. *Immunity* 2017;47:1083–99.
- Le DT, Durham JN, Smith KN, *et al*. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 2017;357:409–13.
- Fancello L, Gandini S, Pelicci PG, *et al*. Tumor mutational burden quantification from targeted gene panels: major advancements and challenges. *J Immunother Cancer* 2019;7:183.
- Ayers M, Lunceford J, Nebozhyn M, *et al*. IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* 2017;127:2930–40.

- 44 Van Allen EM, Miao D, Schilling B, *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 2015;350:207–11.
- 45 Hugo W, Zaretsky JM, Sun L, *et al.* Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* 2016;165:35–44.
- 46 Ferrucci PF, Ascierto PA, Pigozzo J, *et al.* Baseline neutrophils and derived neutrophil-to-lymphocyte ratio: prognostic relevance in metastatic melanoma patients receiving ipilimumab. *Ann Oncol* 2016;27:732–8.
- 47 Diem S, Kasenda B, Spain L, *et al.* Serum lactate dehydrogenase as an early marker for outcome in patients treated with anti-PD-1 therapy in metastatic melanoma. *Br J Cancer* 2016;114:256–61.
- 48 Lim SY, Lee JH, Gide TN, *et al.* Circulating cytokines predict immune-related toxicity in melanoma patients receiving Anti-PD-1-Based immunotherapy. *Clin Cancer Res* 2019;25:1557–63.
- 49 Teng MWL, Ngiow SF, Ribas A, *et al.* Classifying cancers based on T-cell infiltration and PD-L1. *Cancer Res* 2015;75:2139–45.
- 50 Gordon SR, Maute RL, Dulken BW, *et al.* PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. *Nature* 2017;545:495–9.
- 51 Roh W, Chen P-L, Reuben A, *et al.* Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* 2017;9:eaah3560.
- 52 Hopkins AC, Yarchoan M, Durham JN, *et al.* T cell receptor repertoire features associated with survival in immunotherapy-treated pancreatic ductal adenocarcinoma. *JCI Insight* 2018;3. doi:10.1172/jci.insight.122092. [Epub ahead of print: 12 Jul 2018].
- 53 Routy B, Le Chatelier E, Derosa L, *et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* 2018;359:91–7.
- 54 Gopalakrishnan V, Spencer CN, Nezi L, *et al.* Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* 2018;359:97–103.
- 55 Fessler J, Matson V, Gajewski TF. Exploring the emerging role of the microbiome in cancer immunotherapy. *J Immunother Cancer* 2019;7:108.
- 56 Carbone DP, Reck M, Paz-Ares L, *et al.* First-Line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N Engl J Med* 2017;376:2415–26.
- 57 Reck M, Rodríguez-Abreu D, Robinson AG, *et al.* Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med* 2016;375:1823–33.
- 58 Blank CU, Haanen JB, Ribas A, *et al.* The “cancer immunogram”. *Science* 2016;352:658 LP–60.
- 59 Parikh AR, Leshchiner I, Elagina L, *et al.* Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med* 2019;25:1415–21.
- 60 Menzies AM, Long GV. Optimum dosing of ipilimumab in melanoma: too little, too late? *Lancet Oncol* 2017;18:558–9.
- 61 Garon EB, Hellmann MD, Rizvi NA, *et al.* Five-year overall survival for patients with advanced Non-Small-Cell lung cancer treated with pembrolizumab: results from the phase I KEYNOTE-001 study. *J Clin Oncol* 2019;37:2518–27.
- 62 Melloni GEM, Guida A, Curigliano G, *et al.* Precision trial Drawer, a computational tool to assist planning of Genomics-Driven trials in oncology. *JCO Precis Oncol* 2018;1–16.
- 63 Boyiadzis MM, Kirkwood JM, Marshall JL, *et al.* Significance and implications of FDA approval of pembrolizumab for biomarker-defined disease. *J Immunother Cancer* 2018;6:35.
- 64 Weber J, Mandala M, Del Vecchio M, *et al.* Adjuvant nivolumab versus ipilimumab in resected stage III or IV melanoma. *N Engl J Med* 2017;377:1824–35.
- 65 Huang AC, Orlowski RJ, Xu X, *et al.* A single dose of neoadjuvant PD-1 blockade predicts clinical outcomes in resectable melanoma. *Nat Med* 2019;25:454–61.
- 66 Forde PM, Chaft JE, Smith KN, *et al.* Neoadjuvant PD-1 blockade in resectable lung cancer. *N Engl J Med* 2018;378:1976–86.
- 67 Simon R. Critical review of umbrella, basket, and platform designs for oncology clinical trials. *Clin Pharmacol Ther* 2017;102:934–41.
- 68 Wolf DM, Yau C, Sanil A, *et al.* DNA repair deficiency biomarkers and the 70-gene ultra-high risk signature as predictors of veliparib/carboplatin response in the I-SPY 2 breast cancer trial. *NPJ Breast Cancer* 2017;3:31.
- 69 Flaherty KT, Gray R, Chen A, *et al.* The molecular analysis for therapy choice (NCI-MATCH) trial: lessons for genomic trial design. *J Natl Cancer Inst* 2020. doi:10.1093/jnci/djz245. [Epub ahead of print: 10 Jan 2020].
- 70 Thorlund K, Haggstrom J, Park JJ, *et al.* Key design considerations for adaptive clinical trials: a primer for clinicians. *BMJ* 2018;360:k698.
- 71 Hobbs BP, Chen N, Lee JJ. Controlled multi-arm platform design using predictive probability. *Stat Methods Med Res* 2018;27:65–78.
- 72 Zang Y, Lee JJ. A robust two-stage design identifying the optimal biological dose for phase I/II clinical trials. *Stat Med* 2017;36:27–42.
- 73 Riviere M-K, Yuan Y, Jourdan J-H, *et al.* Phase I/II dose-finding design for molecularly targeted agent: plateau determination using adaptive randomization. *Stat Methods Med Res* 2018;27:466–79.
- 74 Wages NA, Tait C. Seamless phase I/II adaptive design for oncology trials of molecularly targeted agents. *J Biopharm Stat* 2015;25:903–20.
- 75 Chiuhan C, Garrett-Mayer E, Nishimura M. An adaptive dose-finding design based on both safety and immunologic responses in cancer clinical trials. *Stat Biopharm Res* 2018;10:185–95.
- 76 Zhang W, Sargent DJ, Mandrekar S. An adaptive dose-finding design incorporating both toxicity and efficacy. *Stat Med* 2006;25:2365–83.
- 77 Koopmeiners JS, Modiano J. A Bayesian adaptive phase I-II clinical trial for evaluating efficacy and toxicity with delayed outcomes. *Clin Trials* 2014;11:38–48.
- 78 Altzerinakou M-A, Paoletti X. An adaptive design for the identification of the optimal dose using joint modeling of continuous repeated biomarker measurements and time-to-toxicity in phase I/II clinical trials in oncology. *Stat Methods Med Res* 2020;29:508–21.
- 79 Hwang TJ, Franklin JM, Chen CT, *et al.* Efficacy, safety, and regulatory approval of food and drug Administration-Designated breakthrough and Nonbreakthrough cancer medicines. *J Clin Oncol* 2018;36:1805–12.
- 80 US Food And Drug Administration. Complex innovative trial designs pilot program; 2018.
- 81 Shan G, Wilding GE, Hutson AD, *et al.* Optimal adaptive two-stage designs for early phase II clinical trials. *Stat Med* 2016;35:1257–66.
- 82 Schilsky RL. Access versus evidence: the regulators' dilemma. *Clin Trials* 2018;15:240–2.
- 83 Hey SP, Kimmelman J. Are outcome-adaptive allocation trials ethical? *Clin Trials* 2015;12:102–6.
- 84 Simonsen KL, Fracasso PM, Bernstein SH, *et al.* The fast real-time assessment of combination therapies in Immuno-ONcology (FRACTION) program: innovative, high-throughput clinical screening of immunotherapies. *Eur J Cancer* 2018;103:259–66.
- 85 Chau I, Haag GM, Rahma OE, *et al.* MORPHEUS: A phase Ib/II umbrella study platform evaluating the safety and efficacy of multiple cancer immunotherapy (CIT)-based combinations in different tumour types. *Ann Oncol* 2018;29(Suppl 8):viii439–40.
- 86 Ribas A, Chow LQ, Boyd JK, *et al.* Avelumab (MSB0010718C; anti-PD-L1) in combination with other cancer immunotherapies in patients with advanced malignancies: the phase 1b/2 javelin Medley study. *J Clin Oncol* 2016;34:TPS3106.
- 87 Johnson ML, Solomon BJ, Awad MM, *et al.* MORPHEUS: A phase Ib/II multi-trial platform evaluating the safety and efficacy of cancer immunotherapy (CIT)-based combinations in patients (pts) with non-small cell lung cancer (NSCLC). *J Clin Oncol* 2018;36:TPS9105.
- 88 Korn EL, Freidlin B. Adaptive clinical trials: advantages and disadvantages of various adaptive design elements. *J Natl Cancer Inst* 2017;109:djx013.
- 89 Loibl S, O'Shaughnessy J, Untch M, *et al.* Addition of the PARP inhibitor veliparib plus carboplatin or carboplatin alone to standard neoadjuvant chemotherapy in triple-negative breast cancer (BrightNess): a randomised, phase 3 trial. *Lancet Oncol* 2018;19:497–509.