

Prospective Evaluation of Genomics Guided Off-Label Treatment

Corresponding Author: Professor Emile Voest

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)

The authors describe a multi arm, phase II trial evaluating drugs already approved for at least one cancer indication in a broader population of patients who had tumor types not covered by initial approvals. It is not clear from the methods section What treatments met the definition of “standard-of-care treatment options”. Were these only therapies known to improve overall survival? Many guidelines count therapies as standard based on lower levels of evidence and, in some cases, very low levels of evidence. Requiring patients to have received all available therapies would significantly by the ability to detect clinical activity for therapies in an investigational context. Why was whole genome sequencing used for treatment selection when widely available gene panel testing is specifically designed to capture alterations related to approved therapies? Given that most patients will discontinue therapy prior to a first response evaluation due to disease progression, those patients should be reported as part of an intention-to-treat analysis. Were radiographic assessments reviewed centrally? Were patients excluded if they had a tumor type/genotype combination for which the therapy to be offered was already FDA/EMA approved (eg BRAF mutated melanoma)?

What criteria were used in selecting 313 cohorts? There are not that many approved therapies in biomarker defined populations to reach that cohort size. As a specific example, the authors mention a cohort of patients With CDK4/6 alterations treated with palbociclib or ribociclib. At the time this study was conducted and to this day, those drugs are approved for use in combination with hormonal therapy in hormone receptor positive breast cancer, regardless of genetic alterations in CDK4, CDK6, cyclin D or CDKN2A. As such, this example would not constitute extension of therapeutic application by tumor type, when holding biomarker selection constant. In other words, even in breast cancer, this approach would be considered off label.

It is not clear what the start and end dates were for individual cohorts in relation to regulatory approvals for some of the more commonly investigated strategies described here. For example, the FDA approved PD1 antibody use for MSI-high tumors regardless of tumor type in 2017 and for TMB-high tumors in 2020. For BRAF/MEK combination therapy, approvals beyond melanoma were realized during the conduct of this study as well, including a tumor agnostic approval in 2022. In other words, it's not clear what the purpose of this trial was in terms of contributing evidence beyond what was already being generated in specific drug development programs and other multi tumor type, platform trials.

The validity of the conclusions made in this paper are heavily dependent on answers to the questions above. On the one hand, it is self-evident that evidence needs to be generated in support of using cancer therapies approved for one common narrowly defined indication more broadly. And, clinical trials or “structured initiatives” such as the one described here are clearly worthwhile as a means to generate that evidence. One cannot draw conclusions regarding the relevance of this strategy in 2025 given the continued evolution of molecular classification and the investigational therapeutic armamentarium. On the other hand, this trial largely corroborated the results of clinical trials being conducted by commercial sponsors and publicly funded groups with these same therapies at the same time. So, it's not clear what statements can be made regarding the utility or futility of this approach in relation to those contemporaneous efforts.

Referee #2

(Remarks to the Author)

Doctor Verkerk and colleagues present the results from the 1st 16110 patients treated in the DRUP (Drug rediscovery Protocol). This program is a true Tour de force that is influenced precision oncology programs in multiple other countries. The authors present a thoughtful analysis of the program results with important implications for implementation of precision oncology. Especially of note, it highlights the importance of offering off-label treatments in the context of clinical trials so that efficacy or lack thereof can be tracked along with potential toxicities.

Major comments:

- The authors present the results from stage 1 and 2 analysis Across multiple tumor drug cohorts. The authors exclude stage 3 expansions that occurred in the context of successful stage 1 and stage 2 cohorts. 247 patients or 15.2% of the included patients were non evaluable mostly due to insufficient treatment cycles. Presuming that this was it was the study design not to consider these patients evaluable if they came off for clinical progression, please discuss the implications of this and the in the discussion as it may be overestimating potential benefit.
- The author's highlight the toxicity benefit ratio as 438 patients had grade 3 events and 58 patients had grade 4 events with 11 grade 5 events. Was there different toxicity rates in different arms? - this may be important as higher toxicity maybe acceptable in arms where greater therapeutic benefit was observed such as with immunotherapy arms.
- How was TMB-H vs TML-H vs MSI-H defined? Was MSI available on all patients and TMB-H patients TMB-H but not MSI-H?
- Table S5: what are the adverse event: "investigations"?
- Were the MET mutations in exceptional responders MET exon 14 skipping?
- For exceptional responders please share specific mutations

Minor comments

- Figure 1B please consider making amplification and mutation a bit more distinct colors.
- Why was there lower efficacy in CRC for TML-H?
- Please clarify was the HRR deficiency cohort DDR genomic alterations outside of BRCA or HRD by another genomic assay?
- Is BRCA loss a loss of function mutation or BRCA deletion

Congrats to the team,

Funda Meric-Bernstam, MD

Referee #3

(Remarks to the Author)

In their manuscript, Verkerk et al explore the efficacy of drugs matched to genomic alterations in a large nationwide program (DRUP).

This is a very large program, well known in the field and the community expects to see the results. The investigators should be congratulated for this outstanding effort.

With the exception of anti-PD1 Ab, the results are disappointing overall, and this leads to the conclusion by the authors that off-label prescriptions should not be pursued outside of clinical trials.

As always with precision medicine trials that take very long time to deliver, the drugs are outdated for most of them (and sometimes not really matched to the molecular alteration) and some biomarkers are very weak (eg CDK4 pathway). The reviewer would therefore discourage focusing only on the message that off-label prescription should be banned. This is maybe true for a period of oncology ranged from 2016 and 2024, but will maybe not be true in 10 years from now when we will have better knowledge of biology, better clinical decision supporting system (LLM + knowledge cards) and better drugs. Also, the authors did not optimally select the patients, as they acknowledge in the part of the discussion related to ESCAT, OncoKB.

If the authors want to keep this message on off label use of drug as main point of the paper, they should be more cautious. Indeed, the paper suffers from two major limitations (old drug and suboptimal biomarkers) that do not allow to generate a firm conclusion about lack of efficacy of « off label use » of drug. One could argue that with better drugs and a better selection of genomic alterations, the results could have been different.

Instead, the reviewer would encourage the authors to find some positive signal in their large dataset. As example, a large number of patients with primary brain tumor and BRAF mutation are outlier responders. This cohort should be described in the result section.

Minor comments :

« while BRAF inhibitors achieve high response rates in melanoma, they show minimal efficacy in patients with colorectal cancer (CRC) harboring the same mutation » : this is not correct. BRAF inhibitors are effective in patients with CRC.

« The SHIVA trial further emphasized these limitations by failing to show improved outcomes from a (limited) range of off label targeted therapies compared to standard treatment in a tumor-agnostic, randomized setting, along with reporting higher toxicity in the experimental group ». In shiva, most of the drugs were not matched to the molecular alteration or the

alterations were known to be poorly predictive (ER and AR in non breast and non prostate cancer).

Result section refers too much to figures. It would be nice if the reader could have more data to read in the text itself. As example, I would expand the part related to outlier responders and the part « clinical outcome ».

Figure 3a : the match drug-genomic alteration should be reported instead of the genomic alterations alone. What is important for the reader is to check that the drug was indeed matched to the alteration and that the drug was optimally bioactive on the target.

I would be cautious when comparing results based on tissue of origin with PFS or CBR as an endpoint. PFS and CBR have important variations between tumor types. As example, a lot of panels of the figure show that patients with prostate cancer, a more indolent disease, have a better CBR.

PFS curves with IO are good, but it would be worth having the survival curves of patients treated with IO without the genomic alteration in order to know if the long PFS is because of the biomarker itself. Maybe this could be found in a real world database.

There are a lot of responders in the brain / braf cohort. Maybe it would be worth having more information about this cohort in the text and figures.

Version 3:

Reviewer comments:

Referee #1

(Remarks to the Author)

The authors' responses to the prior critiques are clear and the modifications to the manuscript appropriate.

Referee #2

(Remarks to the Author)

The authors have addressed reviewer comments appropriately.

Referee #3

(Remarks to the Author)

the authors have addressed my comments

permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Prof. dr. Emile E. Voest, MD PhD
Department of Medical Oncology, Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands
Email: e.voest@nki.nl

Victoria Aranda, Ph.D.
Deputy Editor, Biomedicine, Genetics & Biotechnology
Nature

Amsterdam, December 3rd, 2025

Dear Dr. Aranda, dear Victoria,

Please find enclosed our detailed responses to the reviewers' comments regarding our manuscript "Hope, Hype, or Harm? Rethinking Off-Label Targeted Therapy Using Data from 1,610 Patients" (2025-07-18255A).

We thank the reviewers sincerely for their thoughtful, constructive, and rigorous evaluations of our work. Their feedback has helped us to substantially improve the clarity, balance, and overall quality of the manuscript. We have addressed each comment carefully and revised the manuscript accordingly.

We hope that our responses and the updated version of the manuscript satisfactorily resolve the points raised and demonstrate our commitment to strengthening the scientific contribution of this study.

Thank you once again for your time and consideration.

Yours sincerely,

Karlijn Verkerk, MD
Emile Voest, MD, PhD

Referee #1 (Remarks to the Author):

The authors describe a multi arm, phase II trial evaluating drugs already approved for at least one cancer indication in a broader population of patients who had tumor types not covered by initial approvals.

1. It is not clear from the methods section What treatments met the definition of “standard-of-care treatment options”. Were these only therapies known to improve overall survival? Many guidelines count therapies as standard based on lower levels of evidence and, in some cases, very low levels of evidence. Requiring patients to have received all available therapies would significantly by the ability to detect clinical activity for therapies in an investigational context.

Response

We thank the reviewer for this important question and apologize that our previous description in the Methods section did not provide sufficient clarity. Patients were required to have exhausted all standard-of-care treatment options as defined by the Dutch Society for Medical Oncology guidelines at the time of trial inclusion. When a patient had a clear contraindication to a guideline-recommended therapy, this treatment could be omitted after consensus agreement among the three principal investigators, all professors in medical oncology.

Importantly, DRUP was started to provide access to approved drugs that potentially could benefit patients outside the formal drug label. If a clinical signal of activity was identified this could lead to additional approval, expansion of reimbursement and access to these drugs.

In line with this premise, patients entered DRUP with heterogeneous prior treatment histories. While guideline-based therapies formed the minimum requirement, many patients had received additional lines of treatment, including investigational agents in other clinical trials. This variability reflects the real-world context in which off-label use is considered and it is acknowledged as a limitation in the revised Discussion (see below). Importantly, DRUP systematically collected detailed information on all prior therapies to allow adjustment for these differences wherever possible. Typically, our patients were similar to patients entering phase 1 studies.

We have clarified the eligibility criteria in the revised Methods section as follows (newly added text underlined):

“Eligible patients are adults diagnosed with advanced or metastatic solid tumors who have exhausted all standard-of-care treatment options according to Dutch Society for Medical Oncology guidelines, have no such options available, or have well justified reasons for opting out of guideline-recommended therapies after consensus agreement by the three principal investigators. In addition to the guideline-defined minimum, patients were allowed to have received any number of additional prior therapies, including experimental agents administered in other clinical trials.”

Additionally, we have added the following to the Discussion section of the manuscript:

“Second, although all patients were required to have exhausted standard-of-care treatment options, the level of prior treatment remained heterogeneous. Many patients had received multiple additional lines of therapy, including experimental agents from other trials. This real-world variability reflects the context in which off-label use is typically considered, but may have reduced our ability to detect clinical activity in some settings.”

2. Why was whole genome sequencing used for treatment selection when widely available gene panel testing is specifically designed to capture alterations related to approved therapies?

Response

We thank the reviewer for this question. Whole genome sequencing was not required for treatment selection and was performed only in a small subset of patients. Identification of actionable targets occurred outside the trial as part of routine clinical diagnostics and could involve any molecular test, including targeted gene panels. Whole genome sequencing was conducted post hoc on mandatory pre-treatment biopsies collected within the trial to enable exploratory biomarker analyses, not to guide therapy decisions.

To prevent further confusion, we have clarified this point in the methods section as follows:

“Prior molecular diagnostics of any type, performed outside the trial as part of routine clinical care or another study, must have identified a potentially actionable molecular target.”

3. Given that most patients will discontinue therapy prior to a first response evaluation due to disease progression, those patients should be reported as part of an intention-to-treat analysis.

Response

We thank the reviewer for highlighting this important issue. As DRUP enrolls a last-line population, some patients discontinue therapy within days because of rapid clinical deterioration, despite careful screening. The study protocol therefore required completion of at least one full treatment cycle before formal response evaluation, as shorter exposure would not allow a meaningful opportunity for the drug to exert any effect. All non-evaluable patients were already reported in Table S2 for full transparency.

We agree that, irrespective of protocol requirements, an intention-to-treat analysis that includes all patients who initiated treatment offers valuable additional context, particularly given the frequency of early discontinuation due to disease progression. We have now added this ITT analysis to the Results section:

“Although per protocol NE patients are excluded from the efficacy analyses in DRUP, a supplementary intention-to-treat analysis of all 1468 patients that started treatment in stage 1/2 DRUP cohorts yielded an ORR of 13.2% (95% CI, 11.5-15.1), CBR of 29.3% (95% CI, 27-31.7), median PFS of 2.1 months (95% CI, 2.0-2.6) and median OS of 6.8 months (95% CI, 6.4-7.4).”

4. Were radiographic assessments reviewed centrally?

Response

We thank the reviewer for this question. Radiographic assessments were not reviewed centrally. All evaluations were performed locally by the treating physicians and radiologists according to the criteria prescribed in the protocol for that specific tumor type (RECIST, RANO etc.).

5. Were patients excluded if they had a tumor type/genotype combination for which the therapy to be offered was already FDA/EMA approved (eg BRAF mutated melanoma)?

Response

Indeed, all patients for which the tumor/genotype combinations had an approved therapy were excluded from the trial, as is described in the methods section: “The identified target must be associated with an FDA- or EMA-approved therapy that is not approved for the patient’s tumor type but is available within DRUP.”

6. What criteria were used in selecting 313 cohorts? There are not that many approved therapies in biomarker defined populations to reach that cohort size. As a specific example, the authors mention a cohort of patients with CDK4/6 alterations treated with palbociclib or ribociclib. At the time this study was conducted and to this day, those drugs are approved for use in combination with hormonal therapy in hormone receptor positive breast cancer, regardless of genetic alterations in CDK4, CDK6, cyclin D or CDKN2A. As such, this example would not constitute extension of therapeutic application by tumor type, when holding biomarker selection constant. In other words, even in breast cancer, this approach would be considered off label.

Response

We thank the reviewer for an opportunity to further clarify the criteria. DRUP was designed as an adaptive, umbrella-type trial in which cohorts are defined by the combination of a drug, a biomarker and a tumor type. This structure means that for each biomarker, multiple tumor-specific cohorts can be opened in parallel. In practice, this allows an almost unlimited number of potential biomarker-tumor type cohorts to be activated, which explains how the trial reached 313 cohorts despite the limited number of biomarkers. This is probably best illustrated by Figure 1B which shows a multitude of combinations of rare tumor types and genomic aberrations. This also shows the complexity of genomics guided precision medicine.

We agree that the selection criteria and matching rules may not have been sufficiently clear in the previous version of the manuscript. To address this, we have now added a revised supplementary Table S7, which lists all drugs for which patients were included, along with the associated genomic and tumor-specific inclusion and exclusion criteria. This provides full transparency of the matching criteria.

Regarding the reviewer’s example of CDK4/6 alterations and treatment with palbociclib or ribociclib (which were initially tumor-specific cohorts later merged in collaboration with our independent data monitoring board): we agree that these genetic alterations are not FDA-approved targets for these drugs, and that using CDK4/6 inhibitors based solely on pathway alterations is considered off-label,

even within breast cancer. DRUP was intentionally designed to also allow the testing of hypothetical targets based on emerging biological rationale, even when prior clinical evidence was limited. In hindsight, and as we reflect on in the Discussion, this was not the most effective strategy. Our results clearly show that cohorts based on well-established, high-ESCAT-ranked targets achieved substantially better outcomes, whereas exploratory targets such as CDK4/6 pathway alterations showed limited efficacy.

At the same time, the high submission rates for these cohorts demonstrate strong clinical interest in targeting this pathway at the time that these cohorts were opened. The DRUP experience is therefore informative: it confirms that while such exploratory approaches may seem attractive, in practice evidence-based actionability frameworks should guide target selection to optimize clinical benefit going forward.

- 7. It is not clear what the start and end dates were for individual cohorts in relation to regulatory approvals for some of the more commonly investigated strategies described here. For example, the FDA approved PD1 antibody use for MSI-high tumors regardless of tumor type in 2017 and for TMB-high tumors in 2020. For BRAF/MEK combination therapy, approvals beyond melanoma were realized during the conduct of this study as well, including a tumor agnostic approval in 2022. In other words, it's not clear what the purpose of this trial was in terms of contributing evidence beyond what was already being generated in specific drug development programs and other multi tumor type, platform trials.**

Response

We thank the reviewer for this important point. In the Netherlands, DRUP has been an ongoing nationwide initiative since 2016. Given the long duration of large precision-oncology efforts, it is inevitable that regulatory approvals and external evidence continue to evolve while the trial is running. We recognize that the timing of cohort activation relative to these developments may not have been sufficiently clear in the original manuscript. To address this, we have now added the start dates to supplementary Table S2, which lists all individual patients together with their cohort, allowing readers to see precisely when patients were enrolled in relation to emerging regulatory approvals.

Although some of the drugs studied in DRUP were undergoing development or gaining approval in other settings during the study period, these approvals were not uniformly available in the Netherlands or Europe. A core purpose of DRUP was therefore to provide patients with structured access to potentially beneficial off-label therapies in a consistent, monitored and learning-oriented framework. This approach reduces undocumented off-label prescribing, which otherwise yields no systematic evidence and may expose patients to risk without generating insight. Furthermore, DRUP has contributed to equal access and opportunities for patients through a transparent and easy inclusion for >35 Dutch hospitals participating in DRUP.

An example where the evidence generated by DRUP has been directly relevant to national decision making is that of MSI-high tumors. Although these tumors received an FDA tumor-agnostic approval in 2017, EMA approval remained pending. The DRUP results were instrumental in establishing off-label reimbursement for MSI-high tumors in the Netherlands, as described in the manuscript. Similar

considerations applied to other targets, such as TMB-high, where global evidence accumulated during the study but local access remained limited.

In summary, DRUP was designed to complement ongoing global drug development by providing a structured national platform for access, evaluation and evidence generation. The added timing information in Supplementary Table S2 now makes the relation between cohort activation and external regulatory milestones fully transparent.

8. **The validity of the conclusions made in this paper are heavily dependent on answers to the questions above. On the one hand, it is self-evident that evidence needs to be generated in support of using cancer therapies approved for one common narrowly defined indication more broadly. And, clinical trials or “structured initiatives” such as the one described here are clearly worthwhile as a means to generate that evidence. One cannot draw conclusions regarding the relevance of this strategy in 2025 given the continued evolution of molecular classification and the investigational therapeutic armamentarium. On the other hand, this trial largely corroborated the results of clinical trials being conducted by commercial sponsors and publicly funded groups with these same therapies at the same time. So, it's not clear what statements can be made regarding the utility or futility of this approach in relation to those contemporaneous efforts.**

Response

We thank the reviewer for this thoughtful and nuanced comment. We agree that the continued evolution of molecular classification and the development of new targeted therapies limit the ability of any study to draw conclusions about the future relevance of precision oncology strategies. In revising the manuscript, we have therefore taken care to nuance our conclusions and to clearly frame them within the scientific and therapeutic context of the study period from 2016 to 2024. The revised Discussion now emphasizes that our findings should not be interpreted as a judgment on the long-term utility of off-label targeted therapy, but rather as evidence supporting the need for structured data collection during this phase of rapid progress.

DRUP was not intended to replace or replicate commercial or publicly funded clinical trials. Its purpose was to offer real world and equitable access to off-label targeted therapies for patients who had exhausted all standard options, while generating high quality, prospective and transparent evidence on both benefit and lack of benefit. In many cases DRUP corroborated the results of contemporaneous trials, which supports the validity of a structured off-label evaluation approach in a real world population. At the same time, DRUP provided added value by identifying exceptional responders in rare molecular and tumor combinations, by generating the evidence that led to national reimbursement of one off-label indication, and by systematically reporting negative results to the medical community that would otherwise remain unavailable.

Below the two last paragraphs of the revised manuscript that harbor most of the changes made in response this comment:

“Several limitations of our study should be considered when interpreting these findings. First, many of the drugs evaluated in DRUP were developed before the recent wave of next-generation targeted therapies and immunotherapies, and therefore may not fully reflect the potential of contemporary precision oncology. Likewise, several molecular alterations included in the study are now recognized

as weak or non-actionable biomarkers, which helps explain the modest activity observed in some cohorts. Second, although all patients were required to have exhausted standard-of-care treatment options, the level of prior treatment remained heterogeneous. Many patients had received multiple additional lines of therapy, including experimental agents from other trials. This real-world variability reflects the context in which off-label use is typically considered, but may have reduced our ability to detect clinical activity in some settings. As the field evolves, improved biological understanding, more selective and potent drugs, and more sophisticated clinical decision-support tools may largely enhance patient selection and treatment benefit. Our results should therefore not be interpreted as evidence that off-label precision medicines are inherently ineffective, particularly given the presence of multiple exceptional responders across molecularly defined subgroups. Rather, our experiences from 2016 to 2024 argue for confining off-label use to structured, data-generating frameworks until new advancements prove otherwise to allow for a better informed and safer choice of treatment.

In conclusion, oncologists often treat patients with off-label anticancer treatment based on a scientific rationale and hope for a positive outcome. However, in the absence of a framework to systematically capture outcomes, this practice carries significant risks of toxicity, costs, and inequitable access. Our analysis of over 1,600 patients treated in DRUP highlights both the challenges and the opportunities of off-label use. While overall activity across all indications was limited, we also observed meaningful clinical benefit in defined subgroups and several exceptional responders. These positive signals underscore that effective off-label use is possible, but that it requires robust evidence, careful biomarker selection and prospective evaluation to ensure stepwise drug evaluation towards potential label expansion. Although better drugs and improved decision-support systems may widen the therapeutic window, our experience indicates that for now off-label precision medicines should only be used within clinical trial settings that ensure robust data collection and systematic outcome monitoring."

Referee #2 (Remarks to the Author):

Doctor Verkerk and colleagues present the results from the 1st 16110 patients treated in the DRUP (Drug rediscovery Protocol). This program is a true Tour de force that is influencing precision oncology programs in multiple other countries. The authors present a thoughtful analysis of the program results with important implications for implementation of precision oncology. Especially of note, it highlights the importance of offering off-label treatments in the context of clinical trials so that efficacy or lack thereof can be tracked along with potential toxicities.

Major comments:

1. The authors present the results from stage 1 and 2 analysis across multiple tumor drug cohorts. The authors exclude stage 3 expansions that occurred in the context of successful stage 1 and stage 2 cohorts. 247 patients or 15.2% of the included patients were non evaluable mostly due to insufficient treatment cycles. Presuming that this was it was the study design not to consider these patients evaluable if they came off for clinical progression, please discuss the implications of this and the in the discussion as it may be overestimating potential benefit.

Response

We thank the reviewer for this important observation. As described in the Methods section, exclusion of non-evaluable patients from the efficacy analyses is indeed prespecified in the study protocol: "Patients who fail to complete at least one full treatment cycle (or fewer than two administrations for intravenous medications with cycle lengths below four weeks) are considered NE and replaced in the efficacy and biomarker analyses." This requirement reflects the last-line nature of the DRUP population, in which some patients, despite careful screening, discontinue treatment within days because of rapid clinical deterioration. Such minimal drug exposure does not provide a meaningful window for assessing treatment effect. Since we aimed to find signals of clinical activity in stage 1 and 2, we refrained from an intent-to-treat analysis per cohort.

To ensure transparency, all per-patient information for non-evaluable cases was already provided in Table S2. In response to the reviewer's comment and to address the concern regarding potential overestimation of benefit, we have now also added an intent-to-treat analysis to the Results section:

"Although per protocol NE patients are excluded from the efficacy analyses in DRUP, a supplementary intention-to-treat analysis of all 1468 patients that started treatment in stage 1/2 DRUP cohorts yielded an ORR of 13.2% (95% CI, 11.5-15.1), CBR of 29.3% (95% CI, 27-31.7), median PFS of 2.1 months (95% CI, 2.0-2.6) and median OS of 6.8 months (95% CI, 6.4-7.4)."

2. The author's highlight the toxicity benefit ratio as 438 patients had grade 3 events and 58 patients had grade 4 events with 11 grade 5 events. Was there different toxicity rates in different arms? - this may be important as higher toxicity maybe acceptable in arms where greater therapeutic benefit was observed such as with immunotherapy arms.

Response

We agree that toxicity needs to be interpreted in the context of the expected therapeutic benefit, and that different toxicity-benefit profiles across treatments may influence clinical acceptability. The heterogeneity of DRUP complicates direct comparisons, as both toxicity and clinical benefit varied not only between drugs but also within drug subgroups across different targets and tumor types. Still, to provide greater clarity, we have now added **Figure S3**, which displays for each treatment the proportion of patients experiencing treatment-related grade ≥ 3 AEs alongside the overall CBR observed in all patients treated with that drug. This allows readers to better appreciate the toxicity–benefit relationship across treatments within the study.

Also, DRUP was supported by an international Independent Data Monitoring Committee which convened regularly to address exactly these questions. Based on their toxicity analysis the IDMC had for example advised to reduce Lenvatinib dosing.

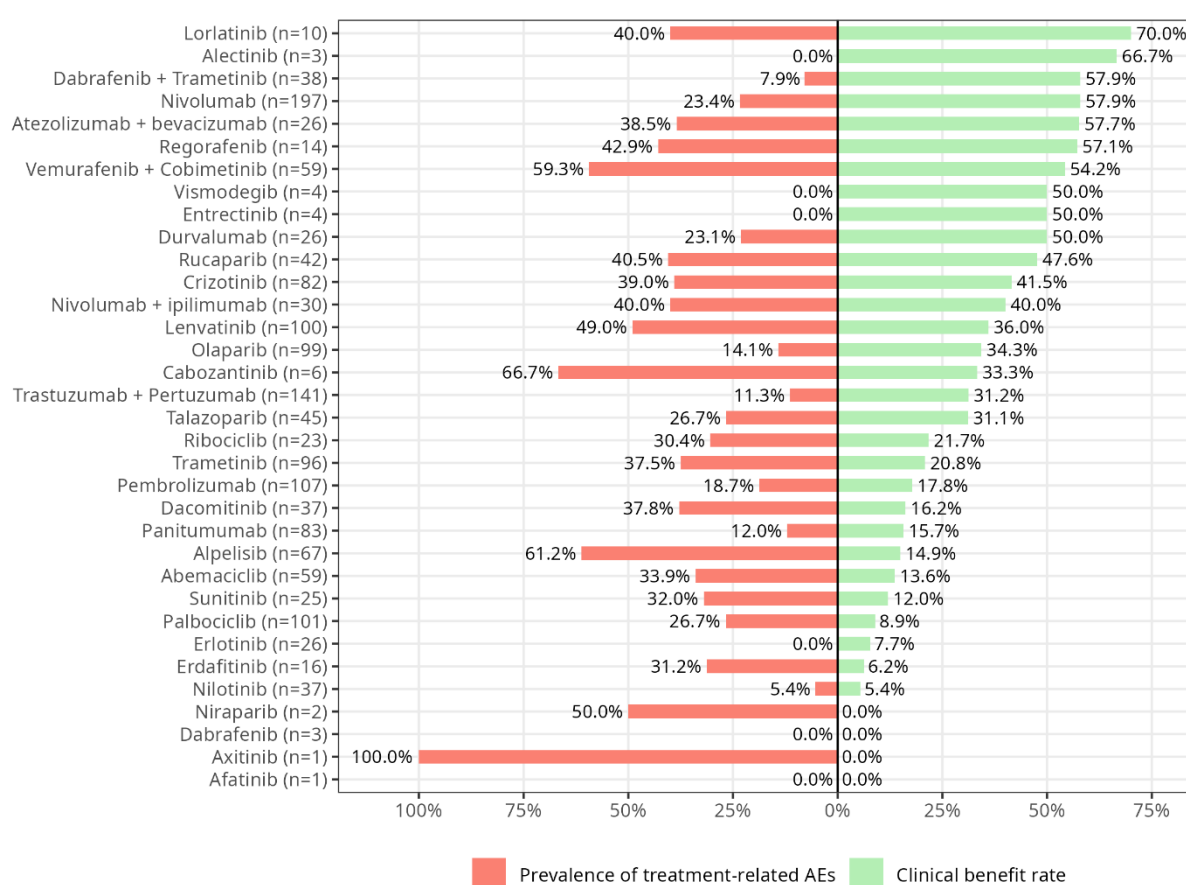


Figure legend: Mirrored bar chart of the clinical benefit rate (in green) and prevalence of treatment-related grade ≥ 3 or higher adverse events (in red) per included treatment. The number of patients that started treatment for each of the respective treatments is given between brackets. Abbreviations: AEs, adverse events.

3. How was TMB-H vs TML-H vs MSI-H defined? Was MSI available on all patients and TMB-H patients TMB-H but not MSI-H?

Response

We thank the reviewer for this opportunity to clarify. In DRUP, it was our intention that all patients enrolled in the TMB-high and TML-high cohorts were microsatellite stable (MSS). This distinction became even more relevant as MSI-H was reimbursed in the Netherlands during the course of the trial. Consequently, MSI testing was required prior to assignment, and patients identified as MSI-H were directed to the MSI-H cohorts/reimbursed treatment rather than to the TMB-H or TML-H cohorts.

However, we acknowledge that in clinical practice, and therefore also in a real-world platform study such as DRUP, perfect separation between these biomarkers is challenging. Testing discordance and differences in sampling time points can lead to occasional misclassification. As we previously reported, some patients initially labelled as MSI-H in routine diagnostics and included as such were later determined to be MSS after comprehensive reassessment (Geurts et al., J Pathol 2024; PMID: 38747304). Conversely, in our recent publication on two TMB-high and TML-high cohorts, some patients classified as MSS in routine testing were found to be MSI-H in the mandatory pre-treatment biopsy (Verkerk et al., Clin Cancer Res 2025; PMID: 41032014).

Thus, while the protocol required MSS status for inclusion in TMB/TML-high cohorts, some patients inevitably “slipped through the cracks,” reflecting the inherent limitations of current diagnostic approaches. Our biomarker analyses, including additional review where appropriate, helped to reduce the impact of such discrepancies on efficacy assessment. Importantly, this variability mirrors real-world clinical practice, where similar discordances may influence treatment decisions.

4. Table S5: what are the adverse event: “investigations”?

Response

We thank the reviewer for highlighting this point. “Investigations” refers to a standardized category in both the Medical Dictionary for Regulatory Activities (MedDRA) and the Common Terminology Criteria for Adverse Events (CTCAE) v4.03, which groups all abnormal laboratory findings under this heading. We recognize that this may not have been immediately clear to readers and have now added a clarifying note to the table legend (Table S6 in the revised manuscript).

5. Were the MET mutations in exceptional responders MET exon 14 skipping? For exceptional responders please share specific mutations

Response

We thank the reviewer for raising this important point. We agree that detailing the specific alterations observed in exceptional responders is essential. These have now been added to the revised Figure 4 as well as directly in the text as described below. As for the MET alterations, two exceptional responders with MET mutations harbored MET exon 14 skipping events, while the last patient carried a MET p.H1094Y tyrosine kinase domain mutation, which was associated with a complete response.

“The targets of the exceptional responders included MSI-H in 21 cases (31.3%), TMB/TML-H in 15 cases (22.4%), 15 BRAF p.V00E mutations (22.4%), four *MET* alterations (6.0%; including two

alterations leading to *MET* exon 14 skipping, one *MET* amplification and one *MET* tyrosine kinase domain mutation [p.H1094Y]), three *ALK* fusions (4.5%), two *BRAC1/2* inactivations (3.0%), two *NRAS* mutations (3.0%; p.G12D and p.Q61R), and one (1.5%) *FGFR2* fusion, *ROS1* fusion and *PIK3CA* p.H10647R mutation.”

Minor comments

- 6. Figure 1B please consider making amplification and mutation a bit more distinct colors.**

Response

We thank the reviewer for this helpful suggestion and have adjusted the color scheme in Figure 1B to make amplifications and mutations more clearly distinguishable, as well as the color schemes in Figures 4 and S2 where similar colors were used.

- 7. Why was there lower efficacy in CRC for TML-H?**

Response

Consistent with prior studies, microsatellite-stable (MSS) CRC proved notoriously difficult to treat with immune checkpoint blockade (ICB). For instance, this is described in the recent systematic review by Guven et al. (*Oncologist* 2024; PMID: 38309719), which reported that across 53 clinical trials the ORR was usually below 10%. This is further supported by other reviews, such as one by El Hajj et al. (*Cancer Treat Rev* 2023; PMID: 37009977), which similarly concludes that MSS CRC shows minimal clinical benefit from ICB due to low immunogenicity, an immune-excluded tumor microenvironment, and lack of predictive biomarkers.

Also, in our recent analysis of two TMB-high/TML-high cohorts treated with ICB (Verkerk et al., *Clin Cancer Res* 2025; PMID: 41032014), none of the MSS CRC patients achieved a response. Our biomarker analysis suggested that the etiology of the TML is a key determinant of response. Tumors with UV-signature or APOBEC-driven mutational processes showed higher response rates, whereas CRC patients in these cohorts predominantly harbored mutations associated with ROS (often related to prior 5-FU exposure) or mutations of unknown etiology, which correlated with lower likelihood of benefit. Additionally, the CRC tumors generally had TMB/TML values at the lower end of the cohort inclusion range – above the threshold for enrollment but not exceptionally high – further contributing to the limited efficacy observed in this subgroup.

- 8. Please clarify was the HRR deficiency cohort DDR genomic alterations outside of BRCA or HRD by another genomic assay?**

Response

For the HRD cohort, a registered HRD test was required for inclusion. This included the CHORD algorithm employed by the Hartwig Medical Foundation, which performs most whole-genome

sequencing in the Netherlands (Nguyen et al., *Nat Commun* 2020; PMID: 33149131). In addition to the HRD cohort, separate cohorts were defined for specific HRR gene alterations outside of BRCA1/2. The full set of options and the matching rules for these cohorts are now detailed in the newly added supplementary Table S7.

9. Is BRCA loss a loss of function mutation or BRCA deletion

Response

We thank the reviewer for this question. In our study, “BRCA loss” refers to biallelic inactivation of the gene, which could result from loss-of-function mutations, deletions, or a combination of both. We have added this clarification to the revised legend of Figure 4, where the term is used.

Referee #3 (Remarks to the Author):

In their manuscript, Verkerk et al explore the efficacy of drugs matched to genomic alterations in a large nationwide program (DRUP).

This is a very large program, well known in the field and the community expects to see the results. The investigators should be congratulated for this outstanding effort.

With the exception of anti-PD1 Ab, the results are disappointing overall, and this leads to the conclusion by the authors that off-label prescriptions should not be pursued outside of clinical trials.

As always with precision medicine trials that take very long time to deliver, the drugs are outdated for most of them (and sometimes not really matched to the molecular alteration) and some biomarkers are very weak (eg CDK4 pathway). The reviewer would therefore discourage focusing only on the message that off-label prescription should be banned. This is maybe true for a period of oncology ranged from 2016 and 2024, but will maybe not be true in 10 years from now when we will have better knowledge of biology, better clinical decision supporting system (LLM + knowledge cards) and better drugs. Also, the authors did not optimally select the patients, as they acknowledge in the part of the discussion related to ESCAT, OncoKB.

If the authors want to keep this message on off label use of drug as main point of the paper, they should be more cautious. Indeed, the paper suffers from two major limitations (old drug and suboptimal biomarkers) that do not allow to generate a firm conclusion about lack of efficacy of « off label use » of drug. One could argue that with better drugs and a better selection of genomic alterations, the results could have been different.

Instead, the reviewer would encourage the authors to find some positive signal in their large dataset. As example, a large number of patients with primary brain tumor and BRAF mutation are outlier responders. This cohort should be described in the result section.

Response

We thank the reviewer for the encouraging comments on the significance of the DRUP program and for the insightful feedback. We agree that the activity of many targeted therapies included in DRUP was modest and that this must be understood in the context of the molecular and therapeutic landscape during the study period from 2016 to 2024. In the revised manuscript we have therefore nuanced our conclusions and made explicit that they apply to this period and are not intended as statements about the long-term prospects of precision oncology.

We acknowledge that several drugs included in DRUP would now be considered outdated and that some biomarkers, such as alterations in the CDK4 pathway, are no longer viewed as strong or actionable. We have integrated these points into a dedicated limitations paragraph, which explains that older drugs, varying biomarker strength and heterogeneity in prior treatments may have contributed to the limited activity observed. These limitations do not allow firm conclusions about how more recent drugs or more selective biomarkers might perform in structured off label settings.

In line with the reviewer's suggestion we have strengthened the positive elements of our findings. The revised Discussion highlights the presence of several exceptional responders across distinct molecular subgroups, including those with BRAF p.V600E mutated primary brain tumors, as an illustration of the potential benefit when the biological rationale is strong. We have also noted that

one cohort generated sufficient evidence to obtain national reimbursement, showing that structured evaluation can meaningfully improve access when efficacy is evident.

Finally, we have clarified that our conclusion is not a call to prohibit off-label prescribing in general. Rather, our experience during the study period supports the use of structured and data generating frameworks as the safest and most informative way to deliver off-label precision therapies at this time. As knowledge, drugs and decision support tools evolve, the role of off-label precision oncology may change.

Having said that, we would also like to emphasize that health insurance companies in the Netherlands received many requests for off-label drug use, which now have been captured and facilitated by DRUP. So, although we agree that several drugs are nearing their patent lifetime (coined “outdated drugs” by this reviewer), they are still commonly used by clinicians without much evidence. DRUP has therefore provided a better, more controlled and simpler access to these targeted anti-cancer drugs and contributed to equality of care because treatment choices did not solely depend on the individual physician.

Below the added/revised paragraphs of the Discussion section of our manuscript are listed:

- Third paragraph: “Despite the overall modest activity across many cohorts, DRUP also demonstrated that substantial benefit is achievable. Several exceptional responders emerged across different molecular subgroups, illustrating the potential of precision oncology to deliver profound and durable benefit, even in rare, heavily pretreated or difficult-to-treat cancers (e.g., 8 exceptional responders with BRAF p.V600E mutated brain tumors). Moreover, one DRUP expansion cohort generated sufficient evidence to achieve full national reimbursement for its off-label indication¹⁹, underscoring that structured evaluation can meaningfully expand patient access when efficacy is clear. These positive signals motivated a deeper examination of the variables in our study design that influence patient benefit, including relevance of molecular targets, number of prior treatments, tissue context, and cohort expansion.”
- Ninth paragraph: “Several limitations of our study should be considered when interpreting these findings. First, many of the drugs evaluated in DRUP were developed before the recent wave of next-generation targeted therapies and immunotherapies, and therefore may not fully reflect the potential of contemporary precision oncology. Likewise, several molecular alterations included in the study are now recognized as weak or non-actionable biomarkers, which helps explain the modest activity observed in some cohorts. Second, although all patients were required to have exhausted standard-of-care treatment options, the level of prior treatment remained heterogeneous. Many patients had received multiple additional lines of therapy, including experimental agents from other trials. This real-world variability reflects the context in which off-label use is typically considered, but may have reduced our ability to detect clinical activity in some settings. As the field evolves, improved biological understanding, more selective and potent drugs, and more sophisticated clinical decision-support tools may largely enhance patient selection and treatment benefit. Our results should therefore not be interpreted as evidence that off-label precision medicines are inherently ineffective, particularly given the presence of multiple exceptional responders across molecularly defined subgroups. Rather, our experiences from 2016 to 2024 argue for confining off-label use to structured, data-generating frameworks until new advancements prove otherwise to allow for a better informed and safer choice of treatment.”
- Tenth/last paragraph: “In conclusion, oncologists often treat patients with off-label anticancer treatment based on a scientific rationale and hope for a positive outcome.

However, in the absence of a framework to systematically capture outcomes, this practice carries significant risks of toxicity, costs, and inequitable access. Our analysis of over 1,600 patients treated in DRUP highlights both the challenges and the opportunities of off-label use. While overall activity across all indications was limited, we also observed meaningful clinical benefit in defined subgroups and several exceptional responders. These positive signals underscore that effective off-label use is possible, but that it requires robust evidence, careful biomarker selection and prospective evaluation to ensure stepwise drug evaluation towards potential label expansion. Although better drugs and improved decision-support systems may widen the therapeutic window, our experience indicates that for now off-label precision medicines should only be used within clinical trial settings that ensure robust data collection and systematic outcome monitoring.

We thank the reviewer for these constructive comments, which have helped us improve the clarity and balance of the manuscript.

Minor comments :

- 1. « while BRAF inhibitors achieve high response rates in melanoma, they show minimal efficacy in patients with colorectal cancer (CRC) harboring the same mutation » : this is not correct. BRAF inhibitors are effective in patients with CRC.**

Response

We thank the reviewer for this comment. Indeed, BRAF inhibitors can be effective in CRC patients when combined with EGFR inhibitors, as demonstrated in the BEACON trial (Kopetz et al., N Engl J Med 2019; PMID: 31566309). Our original statement referred specifically to BRAF inhibitor monotherapy, which is highly effective in melanoma but shows minimal efficacy in CRC without EGFR co-inhibition. We have clarified this in the revised manuscript:

“For example, while *BRAF* inhibitors achieve high response rates in melanoma, they show minimal efficacy when used as monotherapy in patients with colorectal cancer (CRC) harboring the same mutation.”

- 2. « The SHIVA trial further emphasized these limitations by failing to show improved outcomes from a (limited) range of off label targeted therapies compared to standard treatment in a tumor-agnostic, randomized setting, along with reporting higher toxicity in the experimental group ». In shiva, most of the drugs were not matched to the molecular alteration or the alterations were known to be poorly predictive (ER and AR in non breast and non prostate cancer).**

Response

We thank the reviewer for this comment and agree that the SHIVA trial had methodological shortcomings. We have clarified this in the introduction of our manuscript (newly added part underlined).

“The SHIVA trial further emphasized these limitations by failing to show improved outcomes from a (limited) range of off-label targeted therapies compared to standard treatment in a tumor-agnostic, randomized setting, along with reporting higher toxicity in the experimental group. Even though the trial faced methodological limitations, including targets with low predictive value outside their canonical tumor types or poorly matched to therapies, these findings still underscore the importance of carefully evaluating each new indication to ensure patient benefit and avoid unnecessary harm.”

- 3. Result section refers too much to figures. It would be nice if the reader could have more data to read in the text itself. As example, I would expand the part related to outlier responders and the part « clinical outcome ».**

Response

We thank the reviewer for this valuable suggestion. In the previous version of the manuscript, we limited some of the text to remain under the 4,000-word count. However, we fully agree that the Results should be more self-contained, allowing readers to access key data without repeatedly referring to figures and tables. We have therefore expanded several sections to include more detailed information directly in the text (newly added text is underlined):

- Notably, the extent of prior treatment was significantly associated with a decreased likelihood of both CB and OR ($p < 0.001$), as well as a higher baseline LDH ($p < 0.01$) and larger tumor burden (calculated as the baseline sum of target lesions, $p < 0.01$). A higher Eastern Cooperative Oncology Group (ECOG) performance status was associated with decreased chance of CB, but not OR ($p = 0.019$ vs $p = 0.513$).
- Overall duration of response (DoR), progression-free survival (PFS), and overall survival (OS) were 8.2 months (95% CI, 7.6-8.8), 9.4 months (95% CI, 7.8-13.4) and 3.4 months (95% CI, 2.8-3.5), respectively (Figure 2B-D).
- Although per protocol NE patients are excluded from the efficacy analyses in DRUP, a supplementary intention-to-treat analysis of all 1468 patients that started treatment in stage 1/2 DRUP cohorts yielded an ORR of 13.2% (95% CI, 11.5-15.1), CBR of 29.3% (95% CI, 27-31.7), median PFS of 2.1 months (95% CI, 2.0-2.6) and median OS of 6.8 months (95% CI, 6.4-7.4).
- The targets of the exceptional responders included MSI-H in 21 cases (31.3%), TMB/TML-H in 15 cases (22.4%), 15 BRAF p.V00E mutations (22.4%), four MET alterations (6.0%; including two alterations leading to MET exon 14 skipping, one MET amplification and one MET tyrosine kinase domain mutation [p.H1094Y]), three ALK fusions (4.5%), two BRAC1/2 inactivations (3.0%), two NRAS mutations (3.0%; p.G12D and p.Q61R), and one (1.5%) FGFR2 fusion, ROS1 fusion and PIK3CA p.H10647R mutation.

- 4. Figure 3a : the match drug-genomic alteration should be reported instead of the genomic alterations alone. What is important for the reader is to check that the drug was indeed matched to the alteration and that the drug was optimally bioactive on the target.**

Response

We agree that it is important for readers to appreciate the drug-alteration match. In the previous version of the manuscript, this information was available at the per-patient level in Supplementary Table S2; however, we recognize that checking individual patients is inconvenient. Including drugs directly in Figure 3A would render it unreadable, as multiple drugs were used for some targets (e.g., TML-H: Nivolumab [n=25], Nivolumab + Ipilimumab [n=20], Pembrolizumab [n=93]). To address this, we have added supplementary Table S5 in the revised manuscript, which provides a clear overview of drugs used per target, and this table is now referenced in the revised legend of Figure 3A.

- 5. I would be cautious when comparing results based on tissue of origin with PFS or CBR as an endpoint. PFS and CBR have important variations between tumor types. As example, a lot of panels of the figure show that patients with prostate cancer, a more indolent disease, have a better CBR.**

Response

We thank the reviewer for this important comment. For Figure 3A, the primary purpose of including tumor type as a covariate in our multivariable regression models was not to compare tumor types directly, but rather to accurately assess the effect of molecular targets while accounting for the fact that certain targets were disproportionately represented in tumor types with more or less favorable baseline outcomes. By adjusting for tumor type, we reduce the risk that apparent differences in target efficacy are confounded by the underlying biology or clinical course of specific cancers. Tumor-specific data are presented in the figure primarily for transparency, allowing readers to see the full distribution of outcomes across both targets and tumor types, which were both used as covariates in the model. However, we agree that direct comparisons between tumor types using endpoints such as CBR or PFS can be misleading, given intrinsic differences in disease behavior. To address this, we have added a cautionary statement in the Results section:

“Additionally, tumor subgroup data should be interpreted cautiously, as differences in CBR and PFS can reflect intrinsic variations in disease biology rather than differences in treatment efficacy.”

- 6. PFS curves with IO are good, but it would be worth having the survival curves of patients treated with IO without the genomic alteration in order to know if the long PFS is because of the biomarker itself. Maybe this could be found in a real world database.**

Response

We thank the reviewer for this thoughtful suggestion. We agree that comparing PFS in patients treated with immunotherapy with and without the genomic alteration would provide additional insight into biomarker-specific effects. Large, prospectively collected clinical outcome data sets linked to genomic biomarkers are highly needed and DRUP contributes to this. Ultimately this will facilitate designing in silico control groups and determine whether a biomarker is prognostic or predictive. Nonetheless, we respectfully feel that this analysis falls outside the scope of the current study. Survival curves are shown separately for targeted therapy and ICB because these treatment modalities have distinct survival patterns (e.g., tailing of the curve), and we aimed to provide a transparent view of both subgroups. However, the primary focus of this manuscript is to present the overarching results of DRUP after nine years, including all rare cancers which are very much

underserved, and to derive lessons from the experience as a whole.

- 7. There are a lot of responders in the brain / braf cohort. Maybe it would be worth having more information about this cohort in the text and figures.**

Response

We thank the reviewer for this helpful suggestion. In line with this comment, we have now further highlighted the exceptional responders with BRAF p.V600E mutated primary brain tumors in both the Results section and the Discussion (“Several exceptional responders emerged across different molecular subgroups, illustrating the potential of precision oncology to deliver profound and durable benefit, even in rare, heavily pretreated or difficult-to-treat cancers (e.g., 8 exceptional responders with BRAF p.V600E mutated brain tumors).”), and we have placed more emphasis on exceptional responders across the study in general. These additions strengthen the manuscript by illustrating the potential for substantial benefit in specific molecular and tumor contexts.

At the same time, we have not added an in depth description of the entire BRAF brain tumor cohort. Within the current manuscript there is unfortunately no room to expand this cohort in full detail, and after the publication of the ROAR trial (Wen et al, Lancet Oncology 2022), the broader clinical activity of BRAF inhibition in primary brain tumors is already established. Our focus in the present manuscript is therefore on highlighting the most informative and relevant observations from DRUP as a whole.

Prof. dr. Emile E. Voest, MD PhD
Department of Medical Oncology, Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands
Email: e.voest@nki.nl

Victoria Aranda, Ph.D.
Deputy Editor, Biomedicine, Genetics & Biotechnology
Nature

Amsterdam, December 20th, 2025

Dear Dr. Aranda, dear Victoria,

Thank you for reconsidering your initial decision following our appeal. We are grateful for the opportunity to submit a revised version of our manuscript, now titled "*Prospective Evaluation of Genomics Guided Off-Label Treatment*" (2025-07-18255B-Z). We changed the title in line with the suggestions by the reviewers to make it more neutral than our original title "Hype, harm or hope: Rethinking Off-Label Targeted Therapy Using Data from 1,610 Patients".

Our study presents a uniquely large and prospectively collected, comprehensive dataset on the real-world use of off-label targeted therapies across many (often rare) tumor types and molecular alterations. As one reviewer noted, "This program is a true Tour de force that has influenced precision oncology programs in multiple other countries," and another emphasized that "This is a very large program, well known in the field and the community expects to see the results." We appreciate these generous assessments and agree that the size and scope of DRUP allowed us to provide insights that are difficult to obtain from other sources. Patients and physicians always view the use of off-label drugs from a hopeful perspective, but given the toxicity related to these treatments, our study underpins the importance of data collection to be able to make the right decisions for patients. We are proud that *Nature* was the first to publish initial results our study in 2019 and hope that these final analyses will also find its way into *Nature*.

We believe that the revised version of our manuscript fully addresses the concerns raised during peer review and now offers a clearer and more balanced contribution to the field. Based on the reviewers' comments, we have further nuanced both the discussion and the conclusion. The revised manuscript now clearly situates our findings within its scientific and therapeutic landscape and avoids any insufficiently supported forward looking statements about the future of precision oncology. We have highlighted successes with immediate impact in patient access and care. We also emphasized the role of structured data generating frameworks for off-label use during this period of rapid progress. In our opinion, these revisions have strengthened the manuscript and improved its clarity and balance.

We feel that the breadth of the data, the initial positive remarks of two of the reviewers, and the more nuanced interpretation introduced during revision together warrant further consideration for publication in *Nature*. We hope that the revised manuscript meets the journal's standards and offers valuable insights to its readers.

Thank you again for reconsidering our work. We remain grateful for your time and guidance.

Yours sincerely,

Karlijn Verkerk, MD
Emile Voest, MD, PhD

Prof. dr. Emile E. Voest, MD PhD
Department of Medical Oncology, Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands
Email: e.voest@nki.nl

Victoria Aranda, Ph.D.
Deputy Editor, Biomedicine, Genetics & Biotechnology
Nature

Amsterdam, December 20th, 2025

Dear Dr. Aranda, dear Victoria,

Please find enclosed our detailed responses to the reviewers' comments regarding our manuscript, now titled "Prospective Evaluation of Genomics Guided Off-Label Treatment" (2025-07-18255B-Z).

We thank the reviewers sincerely for their thoughtful, constructive, and rigorous evaluations of our work. Their feedback has helped us to substantially improve the clarity, balance, and overall quality of the manuscript. We have addressed each comment carefully and revised the manuscript accordingly.

We hope that our responses and the updated version of the manuscript satisfactorily resolve the points raised and demonstrate our commitment to strengthening the scientific contribution of this study.

Thank you once again for your time and consideration.

Yours sincerely,

Karlijn Verkerk, MD
Emile Voest, MD, PhD

Referee #1 (Remarks to the Author):

The authors describe a multi arm, phase II trial evaluating drugs already approved for at least one cancer indication in a broader population of patients who had tumor types not covered by initial approvals.

1. It is not clear from the methods section What treatments met the definition of “standard-of-care treatment options”. Were these only therapies known to improve overall survival? Many guidelines count therapies as standard based on lower levels of evidence and, in some cases, very low levels of evidence. Requiring patients to have received all available therapies would significantly by the ability to detect clinical activity for therapies in an investigational context.

Response

We thank the reviewer for this important question and apologize that our previous description in the Methods section did not provide sufficient clarity. Patients were required to have exhausted all standard-of-care treatment options as defined by the Dutch Society for Medical Oncology guidelines at the time of trial inclusion. When a patient had a clear contraindication to a guideline-recommended therapy, this treatment could be omitted after consensus agreement among the three principal investigators, all professors in medical oncology.

Importantly, DRUP was started to provide access to approved drugs that potentially could benefit patients outside the formal drug label. If a clinical signal of activity was identified this could lead to additional approval, expansion of reimbursement and access to these drugs.

In line with this premise, patients entered DRUP with heterogeneous prior treatment histories. While guideline-based therapies formed the minimum requirement, many patients had received additional lines of treatment, including investigational agents in other clinical trials. This variability reflects the real-world context in which off-label use is considered and it is acknowledged as a limitation in the revised Discussion (see below). Importantly, DRUP systematically collected detailed information on all prior therapies to allow adjustment for these differences wherever possible. Typically, our patients were similar to patients entering phase 1 studies.

We have clarified the eligibility criteria in the revised Methods section as follows (newly added text underlined):

“Eligible patients are adults diagnosed with advanced or metastatic solid tumors who have exhausted all standard-of-care treatment options according to Dutch Society for Medical Oncology guidelines, have no such options available, or have well justified reasons for opting out of guideline-recommended therapies after consensus agreement by the three principal investigators. In addition to the guideline-defined minimum, patients were allowed to have received any number of additional prior therapies, including experimental agents administered in other clinical trials.”

Additionally, we have added the following to the Discussion section of the manuscript:

“Second, although all patients were required to have exhausted standard-of-care treatment options, the level of prior treatment remained heterogeneous. Many patients had received multiple additional lines of therapy, including experimental agents from other trials. This real-world variability reflects the context in which off-label use is typically considered, but may have reduced our ability to detect clinical activity in some settings.”

2. Why was whole genome sequencing used for treatment selection when widely available gene panel testing is specifically designed to capture alterations related to approved therapies?

Response

We thank the reviewer for this question. Whole genome sequencing was not required for treatment selection and was performed only in a small subset of patients. Identification of actionable targets occurred outside the trial as part of routine clinical diagnostics and could involve any molecular test, including targeted gene panels. Whole genome sequencing was conducted post hoc on mandatory pre-treatment biopsies collected within the trial to enable exploratory biomarker analyses, not to guide therapy decisions.

To prevent further confusion, we have clarified this point in the methods section as follows:

“Prior molecular diagnostics of any type, performed outside the trial as part of routine clinical care or another study, must have identified a potentially actionable molecular target.”

3. Given that most patients will discontinue therapy prior to a first response evaluation due to disease progression, those patients should be reported as part of an intention-to-treat analysis.

Response

We thank the reviewer for highlighting this important issue. As DRUP enrolls a last-line population, some patients discontinue therapy within days because of rapid clinical deterioration, despite careful screening. The study protocol therefore required completion of at least one full treatment cycle before formal response evaluation, as shorter exposure would not allow a meaningful opportunity for the drug to exert any effect. All non-evaluable patients were already reported in Table S2 for full transparency.

We agree that, irrespective of protocol requirements, an intention-to-treat analysis that includes all patients who initiated treatment offers valuable additional context, particularly given the frequency of early discontinuation due to disease progression. We have now added this ITT analysis to the Results section:

“Although per protocol NE patients are excluded from the efficacy analyses in DRUP, a supplementary intention-to-treat analysis of all 1468 patients that started treatment in stage 1/2 DRUP cohorts yielded an ORR of 13.2% (95% CI, 11.5-15.1), CBR of 29.3% (95% CI, 27-31.7), median PFS of 2.1 months (95% CI, 2.0-2.6) and median OS of 6.8 months (95% CI, 6.4-7.4).”

4. Were radiographic assessments reviewed centrally?

Response

We thank the reviewer for this question. Radiographic assessments were not reviewed centrally. All evaluations were performed locally by the treating physicians and radiologists according to the criteria prescribed in the protocol for that specific tumor type (RECIST, RANO etc.).

5. Were patients excluded if they had a tumor type/genotype combination for which the therapy to be offered was already FDA/EMA approved (eg BRAF mutated melanoma)?

Response

Indeed, all patients for which the tumor/genotype combinations had an approved therapy were excluded from the trial, as is described in the methods section: “The identified target must be associated with an FDA- or EMA-approved therapy that is not approved for the patient’s tumor type but is available within DRUP.”

6. What criteria were used in selecting 313 cohorts? There are not that many approved therapies in biomarker defined populations to reach that cohort size. As a specific example, the authors mention a cohort of patients with CDK4/6 alterations treated with palbociclib or ribociclib. At the time this study was conducted and to this day, those drugs are approved for use in combination with hormonal therapy in hormone receptor positive breast cancer, regardless of genetic alterations in CDK4, CDK6, cyclin D or CDKN2A. As such, this example would not constitute extension of therapeutic application by tumor type, when holding biomarker selection constant. In other words, even in breast cancer, this approach would be considered off label.

Response

We thank the reviewer for an opportunity to further clarify the criteria. DRUP was designed as an adaptive, umbrella-type trial in which cohorts are defined by the combination of a drug, a biomarker and a tumor type. This structure means that for each biomarker, multiple tumor-specific cohorts can be opened in parallel. In practice, this allows an almost unlimited number of potential biomarker-tumor type cohorts to be activated, which explains how the trial reached 313 cohorts despite the limited number of biomarkers. This is probably best illustrated by Figure 1B which shows a multitude of combinations of rare tumor types and genomic aberrations. This also shows the complexity of genomics guided precision medicine.

We agree that the selection criteria and matching rules may not have been sufficiently clear in the previous version of the manuscript. To address this, we have now added a revised supplementary Table S7, which lists all drugs for which patients were included, along with the associated genomic and tumor-specific inclusion and exclusion criteria. This provides full transparency of the matching criteria.

Regarding the reviewer’s example of CDK4/6 alterations and treatment with palbociclib or ribociclib (which were initially tumor-specific cohorts later merged in collaboration with our independent data monitoring board): we agree that these genetic alterations are not FDA-approved targets for these drugs, and that using CDK4/6 inhibitors based solely on pathway alterations is considered off-label,

even within breast cancer. DRUP was intentionally designed to also allow the testing of hypothetical targets based on emerging biological rationale, even when prior clinical evidence was limited. In hindsight, and as we reflect on in the Discussion, this was not the most effective strategy. Our results clearly show that cohorts based on well-established, high-ESCAT-ranked targets achieved substantially better outcomes, whereas exploratory targets such as CDK4/6 pathway alterations showed limited efficacy.

At the same time, the high submission rates for these cohorts demonstrate strong clinical interest in targeting this pathway at the time that these cohorts were opened. The DRUP experience is therefore informative: it confirms that while such exploratory approaches may seem attractive, in practice evidence-based actionability frameworks should guide target selection to optimize clinical benefit going forward.

- 7. It is not clear what the start and end dates were for individual cohorts in relation to regulatory approvals for some of the more commonly investigated strategies described here. For example, the FDA approved PD1 antibody use for MSI-high tumors regardless of tumor type in 2017 and for TMB-high tumors in 2020. For BRAF/MEK combination therapy, approvals beyond melanoma were realized during the conduct of this study as well, including a tumor agnostic approval in 2022. In other words, it's not clear what the purpose of this trial was in terms of contributing evidence beyond what was already being generated in specific drug development programs and other multi tumor type, platform trials.**

Response

We thank the reviewer for this important point. In the Netherlands, DRUP has been an ongoing nationwide initiative since 2016. Given the long duration of large precision-oncology efforts, it is inevitable that regulatory approvals and external evidence continue to evolve while the trial is running. We recognize that the timing of cohort activation relative to these developments may not have been sufficiently clear in the original manuscript. To address this, we have now added the start dates to supplementary Table S2, which lists all individual patients together with their cohort, allowing readers to see precisely when patients were enrolled in relation to emerging regulatory approvals.

Although some of the drugs studied in DRUP were undergoing development or gaining approval in other settings during the study period, these approvals were not uniformly available in the Netherlands or Europe. A core purpose of DRUP was therefore to provide patients with structured access to potentially beneficial off-label therapies in a consistent, monitored and learning-oriented framework. This approach reduces undocumented off-label prescribing, which otherwise yields no systematic evidence and may expose patients to risk without generating insight. Furthermore, DRUP has contributed to equal access and opportunities for patients through a transparent and easy inclusion for >35 Dutch hospitals participating in DRUP.

An example where the evidence generated by DRUP has been directly relevant to national decision making is that of MSI-high tumors. Although these tumors received an FDA tumor-agnostic approval in 2017, EMA approval remained pending. The DRUP results were instrumental in establishing off-label reimbursement for MSI-high tumors in the Netherlands, as described in the manuscript. Similar

considerations applied to other targets, such as TMB-high, where global evidence accumulated during the study but local access remained limited.

In summary, DRUP was designed to complement ongoing global drug development by providing a structured national platform for access, evaluation and evidence generation. The added timing information in Supplementary Table S2 now makes the relation between cohort activation and external regulatory milestones fully transparent.

8. **The validity of the conclusions made in this paper are heavily dependent on answers to the questions above. On the one hand, it is self-evident that evidence needs to be generated in support of using cancer therapies approved for one common narrowly defined indication more broadly. And, clinical trials or “structured initiatives” such as the one described here are clearly worthwhile as a means to generate that evidence. One cannot draw conclusions regarding the relevance of this strategy in 2025 given the continued evolution of molecular classification and the investigational therapeutic armamentarium. On the other hand, this trial largely corroborated the results of clinical trials being conducted by commercial sponsors and publicly funded groups with these same therapies at the same time. So, it's not clear what statements can be made regarding the utility or futility of this approach in relation to those contemporaneous efforts.**

Response

We thank the reviewer for this thoughtful and nuanced comment. We agree that the continued evolution of molecular classification and the development of new targeted therapies limit the ability of any study to draw conclusions about the future relevance of precision oncology strategies. In revising the manuscript, we have therefore taken care to nuance our conclusions and to clearly frame them within the scientific and therapeutic context of the study period from 2016 to 2024. The revised Discussion now emphasizes that our findings should not be interpreted as a judgment on the long-term utility of off-label targeted therapy, but rather as evidence supporting the need for structured data collection during this phase of rapid progress.

DRUP was not intended to replace or replicate commercial or publicly funded clinical trials. Its purpose was to offer real world and equitable access to off-label targeted therapies for patients who had exhausted all standard options, while generating high quality, prospective and transparent evidence on both benefit and lack of benefit. In many cases DRUP corroborated the results of contemporaneous trials, which supports the validity of a structured off-label evaluation approach in a real world population. At the same time, DRUP provided added value by identifying exceptional responders in rare molecular and tumor combinations, by generating the evidence that led to national reimbursement of one off-label indication, and by systematically reporting negative results to the medical community that would otherwise remain unavailable.

Below the two last paragraphs of the revised manuscript that harbor most of the changes made in response this comment:

“Several limitations of our study should be considered when interpreting these findings. First, many of the drugs evaluated in DRUP were developed before the recent wave of next-generation targeted therapies and immunotherapies, and therefore may not fully reflect the potential of contemporary precision oncology. Likewise, several molecular alterations included in the study are now recognized

as weak or non-actionable biomarkers, which helps explain the modest activity observed in some cohorts. Second, although all patients were required to have exhausted standard-of-care treatment options, the level of prior treatment remained heterogeneous. Many patients had received multiple additional lines of therapy, including experimental agents from other trials. This real-world variability reflects the context in which off-label use is typically considered, but may have reduced our ability to detect clinical activity in some settings. As the field evolves, improved biological understanding, more selective and potent drugs, and more sophisticated clinical decision-support tools may largely enhance patient selection and treatment benefit. Our results should therefore not be interpreted as evidence that off-label precision medicines are inherently ineffective, particularly given the presence of multiple exceptional responders across molecularly defined subgroups. Rather, our experiences from 2016 to 2024 strongly argue for confining off-label use to structured, data-generating frameworks until new advancements prove otherwise to allow for a better informed and safer choice of treatment.

In conclusion, oncologists often treat patients with off-label anticancer treatment based on a scientific rationale and hope for a positive outcome. However, in the absence of a framework to systematically capture outcomes, this practice carries significant risks of toxicity, costs, and inequitable access. Our analysis of over 1,600 patients treated in DRUP highlights both the challenges and the opportunities of off-label use. While overall activity across all indications was limited, we also observed meaningful clinical benefit in defined subgroups and several exceptional responders. These positive signals underscore that effective off-label use is possible, but that it requires robust evidence, careful biomarker selection and prospective evaluation to ensure stepwise drug evaluation towards potential label expansion. Although better drugs and improved decision-support systems may widen the therapeutic window, our experience indicates that for now off-label precision medicines should only be used within clinical trial settings that ensure robust data collection and systematic outcome monitoring.

Referee #2 (Remarks to the Author):

Doctor Verkerk and colleagues present the results from the 1st 16110 patients treated in the DRUP (Drug rediscovery Protocol). This program is a true Tour de force that is influencing precision oncology programs in multiple other countries. The authors present a thoughtful analysis of the program results with important implications for implementation of precision oncology. Especially of note, it highlights the importance of offering off-label treatments in the context of clinical trials so that efficacy or lack thereof can be tracked along with potential toxicities.

Major comments:

1. The authors present the results from stage 1 and 2 analysis across multiple tumor drug cohorts. The authors exclude stage 3 expansions that occurred in the context of successful stage 1 and stage 2 cohorts. 247 patients or 15.2% of the included patients were non evaluable mostly due to insufficient treatment cycles. Presuming that this was it was the study design not to consider these patients evaluable if they came off for clinical progression, please discuss the implications of this and the in the discussion as it may be overestimating potential benefit.

Response

We thank the reviewer for this important observation. As described in the Methods section, exclusion of non-evaluable patients from the efficacy analyses is indeed prespecified in the study protocol: "Patients who fail to complete at least one full treatment cycle (or fewer than two administrations for intravenous medications with cycle lengths below four weeks) are considered NE and replaced in the efficacy and biomarker analyses." This requirement reflects the last-line nature of the DRUP population, in which some patients, despite careful screening, discontinue treatment within days because of rapid clinical deterioration. Such minimal drug exposure does not provide a meaningful window for assessing treatment effect. Since we aimed to find signals of clinical activity in stage 1 and 2, we refrained from an intent-to-treat analysis per cohort.

To ensure transparency, all per-patient information for non-evaluable cases was already provided in Table S2. In response to the reviewer's comment and to address the concern regarding potential overestimation of benefit, we have now also added an intent-to-treat analysis to the Results section:

"Although per protocol NE patients are excluded from the efficacy analyses in DRUP, a supplementary intention-to-treat analysis of all 1468 patients that started treatment in stage 1/2 DRUP cohorts yielded an ORR of 13.2% (95% CI, 11.5-15.1), CBR of 29.3% (95% CI, 27-31.7), median PFS of 2.1 months (95% CI, 2.0-2.6) and median OS of 6.8 months (95% CI, 6.4-7.4)."

2. The author's highlight the toxicity benefit ratio as 438 patients had grade 3 events and 58 patients had grade 4 events with 11 grade 5 events. Was there different toxicity rates in different arms? - this may be important as higher toxicity maybe acceptable in arms where greater therapeutic benefit was observed such as with immunotherapy arms.

Response

We agree that toxicity needs to be interpreted in the context of the expected therapeutic benefit, and that different toxicity-benefit profiles across treatments may influence clinical acceptability. The heterogeneity of DRUP complicates direct comparisons, as both toxicity and clinical benefit varied not only between drugs but also within drug subgroups across different targets and tumor types. Still, to provide greater clarity, we have now added **Figure S3**, which displays for each treatment the proportion of patients experiencing treatment-related grade ≥ 3 AEs alongside the overall CBR observed in all patients treated with that drug. This allows readers to better appreciate the toxicity–benefit relationship across treatments within the study.

Also, DRUP was supported by an international Independent Data Monitoring Committee which convened regularly to address exactly these questions. Based on their toxicity analysis the IDMC had for example advised to reduce Lenvatinib dosing.

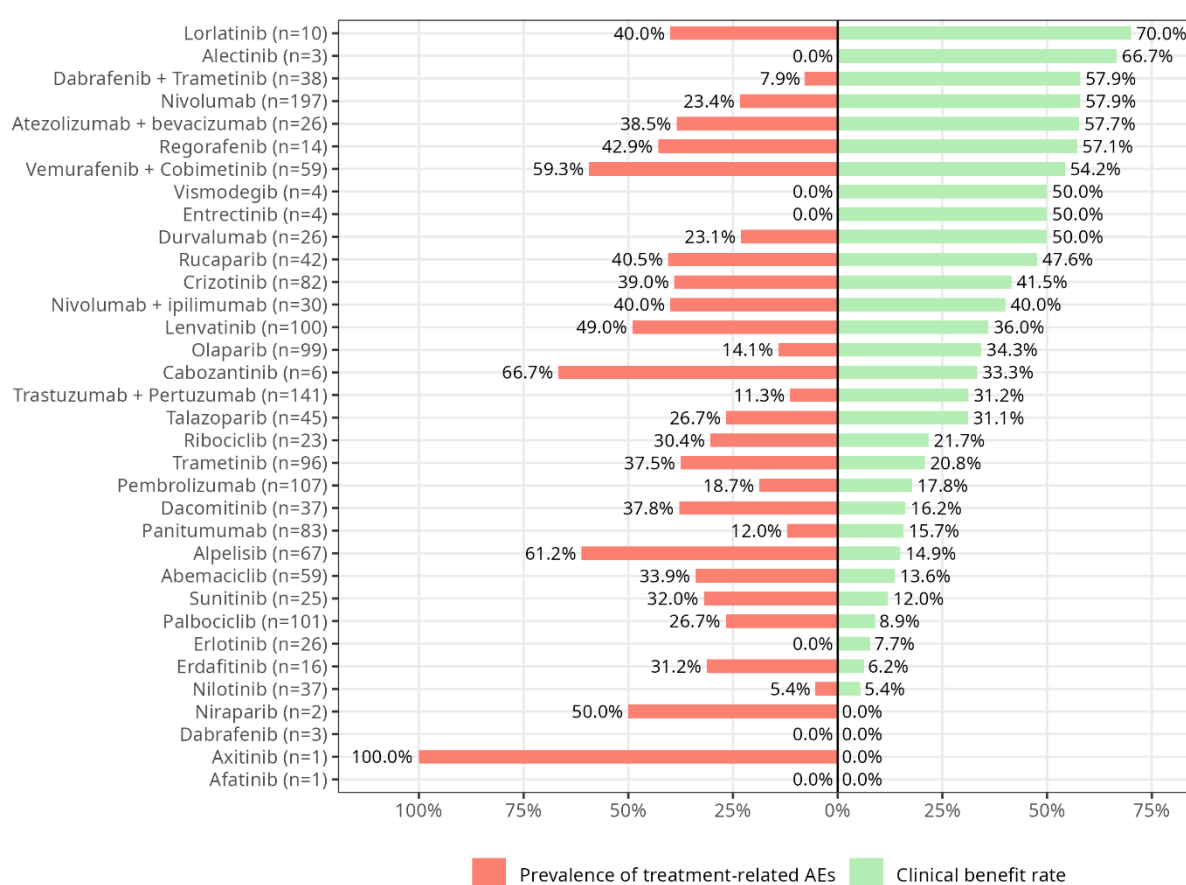


Figure legend: Mirrored bar chart of the clinical benefit rate (in green) and prevalence of treatment-related grade ≥ 3 or higher adverse events (in red) per included treatment. The number of patients that started treatment for each of the respective treatments is given between brackets. Abbreviations: AEs, adverse events.

3. How was TMB-H vs TML-H vs MSI-H defined? Was MSI available on all patients and TMB-H patients TMB-H but not MSI-H?

Response

We thank the reviewer for this opportunity to clarify. In DRUP, it was our intention that all patients enrolled in the TMB-high and TML-high cohorts were microsatellite stable (MSS). This distinction became even more relevant as MSI-H was reimbursed in the Netherlands during the course of the trial. Consequently, MSI testing was required prior to assignment, and patients identified as MSI-H were directed to the MSI-H cohorts/reimbursed treatment rather than to the TMB-H or TML-H cohorts.

However, we acknowledge that in clinical practice, and therefore also in a real-world platform study such as DRUP, perfect separation between these biomarkers is challenging. Testing discordance and differences in sampling time points can lead to occasional misclassification. As we previously reported, some patients initially labelled as MSI-H in routine diagnostics and included as such were later determined to be MSS after comprehensive reassessment (Geurts et al., J Pathol 2024; PMID: 38747304). Conversely, in our recent publication on two TMB-high and TML-high cohorts, some patients classified as MSS in routine testing were found to be MSI-H in the mandatory pre-treatment biopsy (Verkerk et al., Clin Cancer Res 2025; PMID: 41032014).

Thus, while the protocol required MSS status for inclusion in TMB/TML-high cohorts, some patients inevitably “slipped through the cracks,” reflecting the inherent limitations of current diagnostic approaches. Our biomarker analyses, including additional review where appropriate, helped to reduce the impact of such discrepancies on efficacy assessment. Importantly, this variability mirrors real-world clinical practice, where similar discordances may influence treatment decisions.

4. Table S5: what are the adverse event: “investigations”?

Response

We thank the reviewer for highlighting this point. “Investigations” refers to a standardized category in both the Medical Dictionary for Regulatory Activities (MedDRA) and the Common Terminology Criteria for Adverse Events (CTCAE) v4.03, which groups all abnormal laboratory findings under this heading. We recognize that this may not have been immediately clear to readers and have now added a clarifying note to the table legend (Table S6 in the revised manuscript).

5. Were the MET mutations in exceptional responders MET exon 14 skipping? For exceptional responders please share specific mutations

Response

We thank the reviewer for raising this important point. We agree that detailing the specific alterations observed in exceptional responders is essential. These have now been added to the revised Figure 4 as well as directly in the text as described below. As for the MET alterations, two exceptional responders with MET mutations harbored MET exon 14 skipping events, while the last patient carried a MET p.H1094Y tyrosine kinase domain mutation, which was associated with a complete response.

“The targets of the exceptional responders included MSI-H in 21 cases (31.3%), TMB/TML-H in 15 cases (22.4%), 15 BRAF p.V00E mutations (22.4%), four *MET* alterations (6.0%; including two

alterations leading to *MET* exon 14 skipping, one *MET* amplification and one *MET* tyrosine kinase domain mutation [p.H1094Y]), three *ALK* fusions (4.5%), two *BRAC1/2* inactivations (3.0%), two *NRAS* mutations (3.0%; p.G12D and p.Q61R), and one (1.5%) *FGFR2* fusion, *ROS1* fusion and *PIK3CA* p.H10647R mutation.”

Minor comments

- 6. Figure 1B please consider making amplification and mutation a bit more distinct colors.**

Response

We thank the reviewer for this helpful suggestion and have adjusted the color scheme in Figure 1B to make amplifications and mutations more clearly distinguishable, as well as the color schemes in Figures 4 and S2 where similar colors were used.

- 7. Why was there lower efficacy in CRC for TML-H?**

Response

Consistent with prior studies, microsatellite-stable (MSS) CRC proved notoriously difficult to treat with immune checkpoint blockade (ICB). For instance, this is described in the recent systematic review by Guven et al. (*Oncologist* 2024; PMID: 38309719), which reported that across 53 clinical trials the ORR was usually below 10%. This is further supported by other reviews, such as one by El Hajj et al. (*Cancer Treat Rev* 2023; PMID: 37009977), which similarly concludes that MSS CRC shows minimal clinical benefit from ICB due to low immunogenicity, an immune-excluded tumor microenvironment, and lack of predictive biomarkers.

Also, in our recent analysis of two TMB-high/TML-high cohorts treated with ICB (Verkerk et al., *Clin Cancer Res* 2025; PMID: 41032014), none of the MSS CRC patients achieved a response. Our biomarker analysis suggested that the etiology of the TML is a key determinant of response. Tumors with UV-signature or APOBEC-driven mutational processes showed higher response rates, whereas CRC patients in these cohorts predominantly harbored mutations associated with ROS (often related to prior 5-FU exposure) or mutations of unknown etiology, which correlated with lower likelihood of benefit. Additionally, the CRC tumors generally had TMB/TML values at the lower end of the cohort inclusion range – above the threshold for enrollment but not exceptionally high – further contributing to the limited efficacy observed in this subgroup.

- 8. Please clarify was the HRR deficiency cohort DDR genomic alterations outside of BRCA or HRD by another genomic assay?**

Response

For the HRD cohort, a registered HRD test was required for inclusion. This included the CHORD algorithm employed by the Hartwig Medical Foundation, which performs most whole-genome

sequencing in the Netherlands (Nguyen et al., *Nat Commun* 2020; PMID: 33149131). In addition to the HRD cohort, separate cohorts were defined for specific HRR gene alterations outside of BRCA1/2. The full set of options and the matching rules for these cohorts are now detailed in the newly added supplementary Table S7.

9. Is BRCA loss a loss of function mutation or BRCA deletion

Response

We thank the reviewer for this question. In our study, “BRCA loss” refers to biallelic inactivation of the gene, which could result from loss-of-function mutations, deletions, or a combination of both. We have added this clarification to the revised legend of Figure 4, where the term is used.

Referee #3 (Remarks to the Author):

In their manuscript, Verkerk et al explore the efficacy of drugs matched to genomic alterations in a large nationwide program (DRUP).

This is a very large program, well known in the field and the community expects to see the results. The investigators should be congratulated for this outstanding effort.

With the exception of anti-PD1 Ab, the results are disappointing overall, and this leads to the conclusion by the authors that off-label prescriptions should not be pursued outside of clinical trials.

As always with precision medicine trials that take very long time to deliver, the drugs are outdated for most of them (and sometimes not really matched to the molecular alteration) and some biomarkers are very weak (eg CDK4 pathway). The reviewer would therefore discourage focusing only on the message that off-label prescription should be banned. This is maybe true for a period of oncology ranged from 2016 and 2024, but will maybe not be true in 10 years from now when we will have better knowledge of biology, better clinical decision supporting system (LLM + knowledge cards) and better drugs. Also, the authors did not optimally select the patients, as they acknowledge in the part of the discussion related to ESCAT, OncoKB.

If the authors want to keep this message on off label use of drug as main point of the paper, they should be more cautious. Indeed, the paper suffers from two major limitations (old drug and suboptimal biomarkers) that do not allow to generate a firm conclusion about lack of efficacy of « off label use » of drug. One could argue that with better drugs and a better selection of genomic alterations, the results could have been different.

Instead, the reviewer would encourage the authors to find some positive signal in their large dataset. As example, a large number of patients with primary brain tumor and BRAF mutation are outlier responders. This cohort should be described in the result section.

Response

We thank the reviewer for the encouraging comments on the significance of the DRUP program and for the insightful feedback. We agree that the activity of many targeted therapies included in DRUP was modest and that this must be understood in the context of the molecular and therapeutic landscape during the study period from 2016 to 2024. In the revised manuscript we have therefore nuanced our conclusions and made explicit that they apply to this period and are not intended as statements about the long-term prospects of precision oncology.

We acknowledge that several drugs included in DRUP would now be considered outdated and that some biomarkers, such as alterations in the CDK4 pathway, are no longer viewed as strong or actionable. We have integrated these points into a dedicated limitations paragraph, which explains that older drugs, varying biomarker strength and heterogeneity in prior treatments may have contributed to the limited activity observed. These limitations do not allow firm conclusions about how more recent drugs or more selective biomarkers might perform in structured off label settings.

In line with the reviewer's suggestion we have strengthened the positive elements of our findings. The revised Discussion highlights the presence of several exceptional responders across distinct molecular subgroups, including those with BRAF p.V600E mutated primary brain tumors, as an illustration of the potential benefit when the biological rationale is strong. We have also noted that

one cohort generated sufficient evidence to obtain national reimbursement, showing that structured evaluation can meaningfully improve access when efficacy is evident.

Finally, we have clarified that our conclusion is not a call to prohibit off-label prescribing in general. Rather, our experience during the study period supports the use of structured and data generating frameworks as the safest and most informative way to deliver off-label precision therapies at this time. As knowledge, drugs and decision support tools evolve, the role of off-label precision oncology may change.

Having said that, we would also like to emphasize that health insurance companies in the Netherlands received many requests for off-label drug use, which now have been captured and facilitated by DRUP. So, although we agree that several drugs are nearing their patent lifetime (coined “outdated drugs” by this reviewer), they are still commonly used by clinicians without much evidence. DRUP has therefore provided a better, more controlled and simpler access to these targeted anti-cancer drugs and contributed to equality of care because treatment choices did not solely depend on the individual physician.

Below the added/revised paragraphs of the Discussion section of our manuscript are listed:

- Third paragraph: “Despite the overall modest activity across many cohorts, DRUP also demonstrated that substantial benefit is achievable. Several exceptional responders emerged across different molecular subgroups, illustrating the potential of precision oncology to deliver profound and durable benefit, even in rare, heavily pretreated or difficult-to-treat cancers (e.g., 8 exceptional responders with BRAF p.V600E mutated brain tumors). Moreover, one DRUP expansion cohort generated sufficient evidence to achieve full national reimbursement for its off-label indication¹⁹, underscoring that structured evaluation can meaningfully expand patient access when efficacy is clear. These positive signals motivated a deeper examination of the variables in our study design that influence patient benefit, including relevance of molecular targets, number of prior treatments, tissue context, and cohort expansion.”
- Ninth paragraph: “Several limitations of our study should be considered when interpreting these findings. First, many of the drugs evaluated in DRUP were developed before the recent wave of next-generation targeted therapies and immunotherapies, and therefore may not fully reflect the potential of contemporary precision oncology. Likewise, several molecular alterations included in the study are now recognized as weak or non-actionable biomarkers, which helps explain the modest activity observed in some cohorts. Second, although all patients were required to have exhausted standard-of-care treatment options, the level of prior treatment remained heterogeneous. Many patients had received multiple additional lines of therapy, including experimental agents from other trials. This real-world variability reflects the context in which off-label use is typically considered, but may have reduced our ability to detect clinical activity in some settings. As the field evolves, improved biological understanding, more selective and potent drugs, and more sophisticated clinical decision-support tools may largely enhance patient selection and treatment benefit. Our results should therefore not be interpreted as evidence that off-label precision medicines are inherently ineffective, particularly given the presence of multiple exceptional responders across molecularly defined subgroups. Rather, our experiences from 2016 to 2024 strongly argue for confining off-label use to structured, data-generating frameworks until new advancements prove otherwise to allow for a better informed and safer choice of treatment.”

- Tenth/last paragraph: “In conclusion, oncologists often treat patients with off-label anticancer treatment based on a scientific rationale and hope for a positive outcome. However, in the absence of a framework to systematically capture outcomes, this practice carries significant risks of toxicity, costs, and inequitable access. Our analysis of over 1,600 patients treated in DRUP highlights both the challenges and the opportunities of off-label use. While overall activity across all indications was limited, we also observed meaningful clinical benefit in defined subgroups and several exceptional responders. These positive signals underscore that effective off-label use is possible, but that it requires robust evidence, careful biomarker selection and prospective evaluation to ensure stepwise drug evaluation towards potential label expansion. Although better drugs and improved decision-support systems may widen the therapeutic window, our experience indicates that for now off-label precision medicines should only be used within clinical trial settings that ensure robust data collection and systematic outcome monitoring.”

We thank the reviewer for these constructive comments, which have helped us improve the clarity and balance of the manuscript.

Minor comments :

1. « while BRAF inhibitors achieve high response rates in melanoma, they show minimal efficacy in patients with colorectal cancer (CRC) harboring the same mutation » : this is not correct. BRAF inhibitors are effective in patients with CRC.

Response

We thank the reviewer for this comment. Indeed, BRAF inhibitors can be effective in CRC patients when combined with EGFR inhibitors, as demonstrated in the BEACON trial (Kopetz et al., N Engl J Med 2019; PMID: 31566309). Our original statement referred specifically to BRAF inhibitor monotherapy, which is highly effective in melanoma but shows minimal efficacy in CRC without EGFR co-inhibition. We have clarified this in the revised manuscript:

“For example, while *BRAF* inhibitors achieve high response rates in melanoma, they show minimal efficacy when used as monotherapy in patients with colorectal cancer (CRC) harboring the same mutation.”

2. « The SHIVA trial further emphasized these limitations by failing to show improved outcomes from a (limited) range of off label targeted therapies compared to standard treatment in a tumor-agnostic, randomized setting, along with reporting higher toxicity in the experimental group ». In shiva, most of the drugs were not matched to the molecular alteration or the alterations were known to be poorly predictive (ER and AR in non breast and non prostate cancer).

Response

We thank the reviewer for this comment and agree that the SHIVA trial had methodological shortcomings. We have clarified this in the introduction of our manuscript (newly added part underlined).

“The SHIVA trial further emphasized these limitations by failing to show improved outcomes from a (limited) range of off-label targeted therapies compared to standard treatment in a tumor-agnostic, randomized setting, along with reporting higher toxicity in the experimental group. Even though the trial faced methodological limitations, including targets with low predictive value outside their canonical tumor types or poorly matched to therapies, these findings still underscore the importance of carefully evaluating each new indication to ensure patient benefit and avoid unnecessary harm.”

- 3. Result section refers too much to figures. It would be nice if the reader could have more data to read in the text itself. As example, I would expand the part related to outlier responders and the part « clinical outcome ».**

Response

We thank the reviewer for this valuable suggestion. In the previous version of the manuscript, we limited some of the text to remain under the 4,000-word count. However, we fully agree that the Results should be more self-contained, allowing readers to access key data without repeatedly referring to figures and tables. We have therefore expanded several sections to include more detailed information directly in the text (newly added text is underlined):

- Notably, the extent of prior treatment was significantly associated with a decreased likelihood of both CB and OR ($p < 0.001$), as well as a higher baseline LDH ($p < 0.01$) and larger tumor burden (calculated as the baseline sum of target lesions, $p < 0.01$). A higher Eastern Cooperative Oncology Group (ECOG) performance status was associated with decreased chance of CB, but not OR ($p = 0.019$ vs $p = 0.513$).
 - Overall duration of response (DoR), progression-free survival (PFS), and overall survival (OS) were 8.2 months (95% CI, 7.6-8.8), 9.4 months (95% CI, 7.8-13.4) and 3.4 months (95% CI, 2.8-3.5), respectively (Figure 2B-D).
 - Although per protocol NE patients are excluded from the efficacy analyses in DRUP, a supplementary intention-to-treat analysis of all 1468 patients that started treatment in stage 1/2 DRUP cohorts yielded an ORR of 13.2% (95% CI, 11.5-15.1), CBR of 29.3% (95% CI, 27-31.7), median PFS of 2.1 months (95% CI, 2.0-2.6) and median OS of 6.8 months (95% CI, 6.4-7.4).
 - The targets of the exceptional responders included MSI-H in 21 cases (31.3%), TMB/TML-H in 15 cases (22.4%), 15 BRAF p.V00E mutations (22.4%), four *MET* alterations (6.0%; including two alterations leading to *MET* exon 14 skipping, one *MET* amplification and one *MET* tyrosine kinase domain mutation [p.H1094Y]), three *ALK* fusions (4.5%), two *BRAC1/2* inactivations (3.0%), two *NRAS* mutations (3.0%; p.G12D and p.Q61R), and one (1.5%) *FGFR2* fusion, *ROS1* fusion and *PIK3CA* p.H10647R mutation.
- 4. Figure 3a : the match drug-genomic alteration should be reported instead of the genomic alterations alone. What is important for the reader is to check that the drug was indeed matched to the alteration and that the drug was optimally bioactive on the target.**

Response

We agree that it is important for readers to appreciate the drug-alteration match. In the previous version of the manuscript, this information was available at the per-patient level in Supplementary Table S2; however, we recognize that checking individual patients is inconvenient. Including drugs directly in Figure 3A would render it unreadable, as multiple drugs were used for some targets (e.g., TML-H: Nivolumab [n=25], Nivolumab + Ipilimumab [n=20], Pembrolizumab [n=93]). To address this, we have added supplementary Table S5 in the revised manuscript, which provides a clear overview of drugs used per target, and this table is now referenced in the revised legend of Figure 3A.

- 5. I would be cautious when comparing results based on tissue of origin with PFS or CBR as an endpoint. PFS and CBR have important variations between tumor types. As example, a lot of panels of the figure show that patients with prostate cancer, a more indolent disease, have a better CBR.**

Response

We thank the reviewer for this important comment. For Figure 3A, the primary purpose of including tumor type as a covariate in our multivariable regression models was not to compare tumor types directly, but rather to accurately assess the effect of molecular targets while accounting for the fact that certain targets were disproportionately represented in tumor types with more or less favorable baseline outcomes. By adjusting for tumor type, we reduce the risk that apparent differences in target efficacy are confounded by the underlying biology or clinical course of specific cancers. Tumor-specific data are presented in the figure primarily for transparency, allowing readers to see the full distribution of outcomes across both targets and tumor types, which were both used as covariates in the model. However, we agree that direct comparisons between tumor types using endpoints such as CBR or PFS can be misleading, given intrinsic differences in disease behavior. To address this, we have added a cautionary statement in the Results section:

“Additionally, comparison of tumor subgroups should be interpreted cautiously, as differences in CBR and PFS can reflect intrinsic variations in disease biology rather than differences in treatment efficacy.”

- 6. PFS curves with IO are good, but it would be worth having the survival curves of patients treated with IO without the genomic alteration in order to know if the long PFS is because of the biomarker itself. Maybe this could be found in a real world database.**

Response

We thank the reviewer for this thoughtful suggestion. We agree that comparing PFS in patients treated with immunotherapy with and without the genomic alteration would provide additional insight into biomarker-specific effects. Large, prospectively collected clinical outcome data sets linked to genomic biomarkers are highly needed and DRUP contributes to this. Ultimately this will facilitate designing in silico control groups and determine whether a biomarker is prognostic or predictive. Nonetheless, we respectfully feel that this analysis falls outside the scope of the current

study. Survival curves are shown separately for targeted therapy and ICB because these treatment modalities have distinct survival patterns (e.g., tailing of the curve), and we aimed to provide a transparent view of both subgroups. However, the primary focus of this manuscript is to present the overarching results of DRUP after nine years, including all rare cancers which are very much underserved, and to derive lessons from the experience as a whole.

7. There are a lot of responders in the brain / braf cohort. Maybe it would be worth having more information about this cohort in the text and figures.

Response

We thank the reviewer for this helpful suggestion. In line with this comment, we have now further highlighted the exceptional responders with BRAF p.V600E mutated primary brain tumors in both the Results section and the Discussion ("Several exceptional responders emerged across different molecular subgroups, illustrating the potential of precision oncology to deliver profound and durable benefit, even in rare, heavily pretreated or difficult-to-treat cancers (e.g., 8 exceptional responders with BRAF p.V600E mutated brain tumors)."), and we have placed more emphasis on exceptional responders across the study in general. These additions strengthen the manuscript by illustrating the potential for substantial benefit in specific molecular and tumor contexts.

At the same time, we have not added an in depth description of the entire BRAF brain tumor cohort. Within the current manuscript there is unfortunately no room to expand this cohort in full detail, and after the publication of the ROAR trial (Wen et al, Lancet Oncology 2022), the broader clinical activity of BRAF inhibition in primary brain tumors is already established. Our focus in the present manuscript is therefore on highlighting the most informative and relevant observations from DRUP as a whole.