White Paper

Opportunities in Evidence Generation for Tumor-agnostic Targeted Therapies to Better Support Health Technology Assessments in Europe

Dr. Rodrigo Dienstmann, Medical Director, Oncoclínicas Precision Medicine, Oncoclínicas Group
Prof. Fabrice Barlesi, General Director, Gustave Roussy
Prof. James Wason, Professor of Biostatistics, Newcastle University
Prof. Allan Hackshaw, Director, Cancer Research UK & UCL Cancer Trials Centre
Emma Brinkley, Innovation Director, Scientific Services, IQVIA
Clarissa Zebrini, Access Strategy Leader for Tumor Agnostic & Rare Cancers, Roche
Gianluca Rossato, Product Development Oncology, Principal Clinical Scientist, Roche
Laura Perez, Principal Real-World Data Scientist, Roche

Table of contents

Executive summary

Tumor-agnostic oncology therapies — those that treat cancer irrespective of the histological subtype but they have the same biological feature such as a mutation — pose significant assessment challenges for regulators and health technology assessment (HTA) bodies. The factors that cause uncertainty in evidence packages for decision makers in these cases, primarily small heterogeneous populations, are not typically an issue with tumor-specific drugs. Thus, the situation requires a different framework for generating and evaluating evidence.

Here we focus on the key challenges of developing evidence for tumor-agnostic therapies in rare populations. We focus especially on opportunities to use Real-World Data (RWD) as part of the evidence to support HTA decisions. Observations and recommendations are drawn from an expert panel of oncologists, epidemiologists, statisticians, RWD experts, and pharmaceutical industry representatives.

The major aspects of the evidence package for tumor-agnostic therapies that pose challenges for decision makers include:

- Direct comparative efficacy data would be lacking when randomized controlled trials are not feasible for tumor agnostic therapies in rare tumours, as they would take many years to complete (Lozano-Ortega et al. 2019);
- The natural history of the disease might not be well understood, including the uncertainty over the prognostic value of a biomarker
- The efficacy of a new therapy or the standard of care (SoC) differ across tumor types
- The testing and treatment paradigm is evolving rapidly.

The expert panel discusses lessons learned in seeking reimbursement for drugs treating tumors that are neurotrophic tropomyosin-receptor kinase (NTRK) gene fusion positive, as an example, and highlight several generalizable lessons.

- Multiple RWD sources may be needed to understand the natural history as well as the prognostic value of these genomic alterations
- Researchers should use current techniques and develop new ones to reliably adjust for underlying differences in patients identified from RWD and those in the relevant Intended-To-Treat (ITT) Randomized Control Trial (RCT) patient populations (e.g., resulting from different access and timing to Next Generation Sequencing (NGS) testing)
- Flexibility is needed in the real-world study design as multiple approaches may be needed to generate sufficient evidence regarding comparative effectiveness (e.g., use of external comparators)

However, these strategies will only be acceptable to HTAs if there is a paradigm shift in reimbursement negotiations; HTA bodies need to accept evidence of treatment effects from RWD as well as clinical trial data which may be with or without a randomized concurrent control arm.

All the above requires that in the future, sponsors, payers, and other stakeholders (e.g., academia, therapy area experts (TAEs), and other pharmaceutical companies, where possible) collaborate more closely to develop a common framework for evidence generation. This work should include developing more common structures across data sources for RWD and electronic health records (EHR) and more unified approaches to data collection. Eventually this will result in data-sharing platforms that allow partners to leverage commonly created data to bridge evidence gaps, given different perspectives.

Introduction

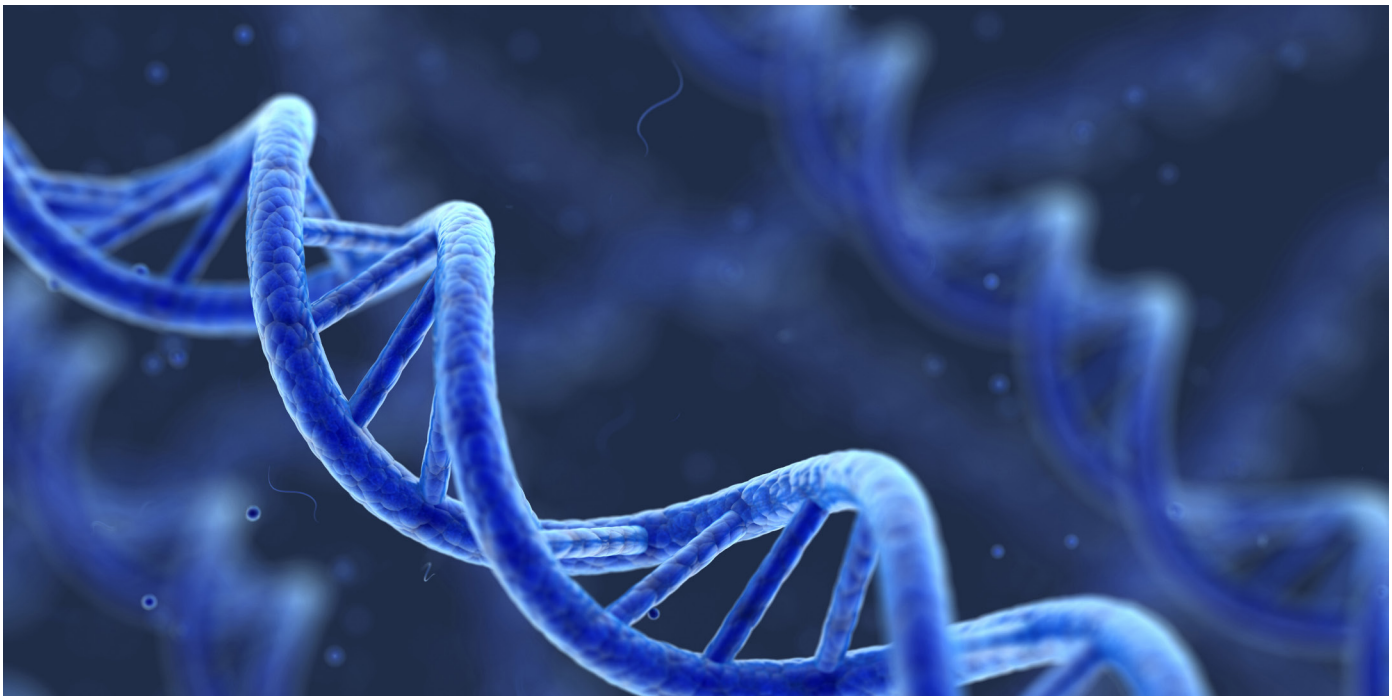
With advances in our understanding of the molecular biology of cancer, treatment paradigms in oncology are rapidly evolving, and innovative therapies are being developed to treat cancer based on specific molecular alterations rather than the traditional tumor (histological) type (Pestana, et al., 2020). Recent regulatory approvals of tumor agnostic/ histology-independent drugs include entrectinib and larotrectinib in Neurotrophic Tropomyosin-Receptor Kinase (NTRK) fusion positive patients, and pembrolizumab in microsatellite instability (MSI-high) and tumor mutational burden (TMB-high) patients. The availability of tumor-agnostic therapies presents new highly effective treatment options for cancer patients, particularly those with advanced disease.

However, tumor-agnostic therapies represent some of the most complex cases to be assessed by regulators and health technology assessment (HTA) bodies (HTx, 2021), and sponsors have experienced challenges and delays in obtaining full approval and reimbursement of these drugs (Grigore et al., 2020). Decision makers face uncertainties related to the inferences they must draw from a small number of patients and/or the heterogeneity of tumor types.

These factors are not typically an issue with tumor-specific drugs, and the situation, therefore, demands a different framework for generating evidence. However, in Europe there is limited guidance or successful examples on how to develop evidence for more successful reimbursement outcomes.

Here, we discuss challenges in, and recommendations for, improving how evidence is generated for tumor agnostic therapies in rare populations. We focus especially on ways to expand the use of Real-World Data (RWD) in generating meaningful comparative evidence to support HTA decisions. The content is drawn from a series of workshops conducted between February 2020 and July 2020 with an expert panel of oncologists, epidemiologists, statisticians, RWD specialists, and pharmaceutical industry representatives.

To illustrate the concepts discussed, we showcase the challenges and opportunities encountered specific to obtaining reimbursement for drugs treating tumors that are NRTK gene fusion positive.



The challenges in determining reimbursement

The major limitations of the evidence framework for tumor agnostic therapies that pose challenges for assessors include the following four issues:

Direct comparative efficacy data would be lacking when randomized controlled trials are not feasible
Not having a concurrent control arm is an important limitation of some trials of tumor-agnostic therapies (Casali, et al., 2015; Hierro, et al., 2019; Pestana, et al., 2020; van Waalwijk,et al., 2019). Therefore, alternative ways of generating comparative effectiveness data are required, and this is likely to come from RWD sources. While RWD and external controls are possible (European Medicines Agency (EMA), 2016), to date, EUnetHTA has not developed guidance as to what approach should be taken in situations where there is no clear single comparator, as is the case with agnostic therapies (EUnetHTA 2015). A multi-disciplinary team of experts have attempted to provide a framework for HTA assessment of histology-independent precision oncology therapies (Gaultney et al., 2021);

Furthermore, efficacy results presented at the time of filing could be considered limited if, for example, the survival data are immature, or there is uncertainty over the clinical relevance of primary outcome measures such as progression-free survival and tumor response rate (HTx 2021).

The natural history of the disease might not be well understood, including the uncertainty over the prognostic value of a biomarker
Characterization of the natural history in rare and heterogenous biomarker-selected cancer patient populations can sometimes be challenging. This is particularly problematic when there is no randomized control arm, so it is not possible to reliably determine whether an observed improvement in efficacy in single arm trials of a targeted agent is due to the drug itself or that the patient's tumor was biomarker positive (or negative). The biomarker is a confounding factor. Also, the prognostic value of the same biomarker may differ between cancer types. The prognostic value of a new biomarker across tumors is often uncertain, with few known examples (e.g. AKT1 in breast cancer (Smyth et al., 2020), and TMB-H or NTRK fusion across several cancer types (Shao et al., 2020; Bazhenova et al., 2021).

Difficulties in evaluating the prognostic value of a biomarker in relation to tumor-agnostic therapies and rare tumors includes: ensuring that all previous anti-cancer treatments have been accounted for, whether the presence of other biomarkers matters, and associated targeted treatments that may confound association.

The efficacy of a new therapy or the standard of care (SoC) differ across tumor types
To date, most cancer trials were based on patients with the same histological cancer type, so their interpretation have been relatively straightforward. However, with tumor-agnostic therapy studies, one issue that has been raised is whether the agent has similar efficacy across the different tumor types. This will be difficult to reliably determine because many specific tumor types will have very few patients (e.g. <10). Statistical methods should therefore be considered that attempt to test that a drug's efficacy is generally homogenous across small patient cohorts. However, there have been few cases where tumor-agnostic drugs were shown to be effective in some cancer types but not others (e.g., BRAF V600 [Hayman, 2015; Kopetz et al., 2015] and high tumor burden (TMB-H) [Sung et al., 2020]). When this has happened, there has usually been a biological explanation.

Another challenge is choosing the SoC control arm if a RCT is possible. Because tumor-agnostic trials consist of multiple tumor types, and even within tumor type patients may be on different lines of therapy, the control arm of these trials are expected to have quite different SoC therapies (e.g. chemotherapies and combination therapies).

The testing and treatment paradigm is evolving rapidly
These issues impact the frequency, acceptability, and interpretation of NGS testing in the context of reimbursement. There is variable access to next generation sequencing (NGS)/advanced diagnostic testing in clinical practice, within and between countries, and this may lead to selected patient populations to be included in an evidence package that may not be generalizable.

Identifying opportunities: The entrectinib NTRK+ case study

BACKGROUND

NTRK fusions are rare, occurring in approximately 0.3% of all solid tumors; however, they can act as oncogenic drivers in a variety of cancer types (Stransky et al., 2014; Okamura et al., 2018). Two drugs have received conditional approval from the European Medicines Agency (EMA) thus far: larotrectinib and entrectinib.

To date, only a few countries in Europe (e.g., Belgium, Finland, and Italy) have recommended these two therapies for reimbursement based on the initial evidence package provided, which demonstrated clinical efficacy from a single-arm trial. Some countries (e.g., England and the Netherlands) have provided conditional funding based on the development of additional evidence. Further negotiations for reimbursement schemes at the country level are ongoing. (Broogard et al., 2022)

In the case of entrectinib, at the time of filing with the EMA (December 2018), the evidence package was based on clinical efficacy from single-arm trials. An RCT for confirmatory data was not considered feasible, which has rendered the reimbursement negotiations for the drug challenging and varied across Europe (Lozano-Ortega et al, 2019).

One of the difficulties in providing additional evidence beyond trial data at the time of filing was the lack of RWD that could capture NTRK+ fusion tumor patients treated by the SoC. Such data were not available, as previously there was no knowledge of the need or incentive to conduct NGS testing to identify these patients. Since then, to our knowledge, only one US study has provided outcome information on NTRK fusion-positive patients, although it included patients who received a Tropomyosin-Receptor Kinase inhibitor (TRKi) and who were not necessarily only in an advanced/metastatic stage (Rosen, 2019).

With the further development of a real-world dataset linking clinical and genomic data at scale, Flatiron Clinico-Genomic database (FH-CGDB), evidence gaps could be filled more directly. FH-CGDB provided information on the natural history of NTRK+ non-entrectinib treated patients in advanced/metastatic solid tumors as well as to evaluate the biomarker for its relationship with its clinical prognosis (See Appendix 1 and 2 for sources and details of the study design and approaches used in this case study.) (Demetri et al, 2021). Although this initial exploratory work was conducted after the filing and reimbursement dossiers, the lessons learned from the NTRK case study are several.

Lessons learned

MULTIPLE RWD SOURCES MAY BE NEEDED TO UNDERSTAND THE NATURAL HISTORY OF PATIENTS WITH RARE INDICATIONS TARGETED BY NEW BIOMARKERS.

There are few data sources that contain relevant and complete data, and those that do, may not contain sufficient numbers of patients.

It is, therefore, critical to develop or have access to data sources that maximize representation of target patient populations, including the trial population. The data sources should also contain the required data elements or variables that allow the heterogeneity of the populations to be described, especially in relation to treatment patterns and outcomes.

The use of different RWD sources will also allow the generalizability/representativeness of the clinical trial population to be assessed qualitatively. Note, though, that data from multiple databases can only be pooled if the data are harmonized across the different sources in terms of patient demographics, medical history, comorbidities, and treatments.

NGS TESTING, WHICH IS NOT YET ROUTINE IN THE STANDARD OF CARE, IS CRITICAL TO THE IDENTIFICATION AND GENERALIZABILITY OF THE TARGET PATIENT POPULATION IN RWD.

All patients do not currently undergo NGS testing to direct an appropriate treatment regimen. Rather, this testing is primarily done only with the most common cancers for which targeted treatments are largely available (i.e., lung, breast, and colorectal cancers) or rare pediatric cancers in the relapse setting.

The main criteria for selecting patients from a RWD source should be the availability of their biomarker status. This is critical to identify comparable patients to those included in a clinical trial setting, and also those who will be treated through routine care following approval. Unfortunately, until testing is universal across tumors and sites, patients identified from RWD may not fully represent the ITT population.

Collecting information on the type of diagnostic tests performed and the definition of the biomarker may also be needed to characterize the patient population. It is important to understand differences in the time of testing in the context of disease progression and treatment decisions.

SUCCESS REQUIRES EARLY PLANNING AND CAREFUL CONSIDERATION OF THE UNDERLYING DATA FOR AVAILABILITY, QUALITY, AND RELIABILITY.

The FH-CGDB analyses used retrospective patient-level data including demographics, diagnostic information (e.g., stage, pathology, molecular information, and radiology), death date, treatments (e.g., line of therapy, dosing, and regimens), and patient outcomes. Missingness of key variables is often a challenge with RWD sources. In the case of the FH-CGBB dataset, some key variables were often missing (i.e., Eastern Cooperative Oncology Group (ECOG)) or unavailable. Also, outcomes relevant for HTA agencies include progression-free survival (PFS) from relevant SoC treatment, and Quality of Life (QoL).

In the post-launch setting, real-world databases may have limitations since the number of untreated patients is unlikely to grow considerably. For instance, it is expected that in the future, all NTRK+ patients in the US will be treated with either larotrectinib or entrectinib. Prospective data collection should be a special consideration to evaluate the effectiveness of new, targeted treatments (e.g., TRKi) in routine care. Prospective registries can still be valid for natural history and prognosis analyses by exclusion or by censoring at the time such treatments are initiated.

Study planning should be done as early as possible – concurrent to the trial design and/or enrollment, if feasible.

RWD CAN ALSO PROVIDE MEANINGFUL EVIDENCE ON THE PROGNOSTIC VALUE OF BIOMARKERS.

Capitalizing on the fact that the FH-CGDB also had information on NTRK- negative tested patients, the Hazard Ratio (HR) for death of fusion-positive vs. matched NTRK biomarker-negative tumor patients was 1.6 (95%CI 1.0-2.5), (Demetri et al., 2021). Interpretation of these results remains difficult as NTRK+ is a relatively new target, and analyses were based on a convenient sample. However, these types of studies can be indirectly supportive evidence of treatment efficacy in a single-arm setting. With an earlier Flatiron dataset in an unselected population, a similar estimate was obtained (HR=1.44; 95%CI, 0.61–3.37; p = 0.648) (Bazhenova, et al., 2021).

In the FH-CGDB analyses, it was possible to match by tumor type, and this was a priority given that the prognosis may vary by tumor type.

It is also important to carefully select the data elements on which to apply the matching methodologies, which at minimum cover ECOG Performance Status (PS), number of previous therapies, histology, and the time between metastatic disease and NGS testing. Data on concurrent molecular alterations is also essential to account for other factors that influence prognosis. Exploratory analyses by age, sex, and race could be done to check that there are no substantial differences in treatment efficacy or adverse events (AEs) to limit the need for matching by these factors.

EXTERNAL CONTROLS FROM RWD CAN GENERATE MEANINGFUL EVIDENCE ON COMPARATIVE EFFECTIVENESS, BUT THERE ARE METHODOLOGICAL CHALLENGES.

Various statistical methods are available for generating comparative effectiveness evidence using external controls, such as propensity score matching or inverse probability of treatment weighting). Methods have been detailed, described, and applied in multiple settings but they typically require relatively large datasets, especially when there are many matching factors (Harder, et al., 2010; Stuart, et al., 2013; Davies, 2018). However, in a tumor agnostic, rare population setting, these types of advanced analyses may not be reliable.

Given the challenge of small patient cohorts, it is recommended, at a minimum, that individual clinical trial and RWD patients be matched, or a weighting methodology be employed, using ECOG, type/stage cancer, and the number of previous lines of therapy (cutoff of >2). While NGS testing is not universally conducted with advanced/metastatic diagnoses to support treatment decisions, the timing of NGS as compared to advanced metastasis and the setting in which treatment is given are also desirable for matching in order to capture the contextual differences of patients between sources.

Other important characteristics thought to impact prognosis should be considered (and may also be included as matching factors) including:

- Treatment naïve (0-1 line) vs. refractory setting (overlapping with prognosis)
- Cancer with an established therapy (e.g., guideline therapy) vs. cancers with no established therapies
- Rare cancer vs. more common cancers
- Cancer with known good prognosis vs. cancer with worse prognosis

These strategies should help improve the reliability of the comparative analyses in rare populations, with the aim of trying to minimize potential bias and confounding. However, a strikingly large treatment effect should make bias less of a problem (Ghadessi, 2020). There are different statistical techniques for addressing unmeasured confounding (such as modelling and simulation on the magnitude of unmeasured confounding) (Sammon et.al., 2020; Thompson, 2021) as well as selection bias (e.g., adjusting for left truncation resulting from delays between diagnosis and molecular testing; Brown et al., 2022).

Key takeaways from the entrectinib case study

1. Identifying potential data sources requires comprehensive data landscaping activities and may entail pooling or linking separate datasets to deliver larger and more representative or complete datasets that account for the heterogeneity of tumors and treatments.
2. Until NGS testing is routinely done at the time of advanced diagnosis, recommendations for data collection include collecting information on the age of tissue samples, sample quality, and bias by institution with respect to the likelihood of the patient receiving genomic testing.
3. Robust study planning — whether for a retrospective real world database study or for prospective primary real world data collection — should be done at a similar time to the clinical trial design. Multiple approaches may be needed to generate sufficient data for HTAs. Moving forward, increasing harmonization of RWD data collection will be required.
4. With thoughtful selection of appropriate data sources paired with the appropriate statistical methods to compare biomarker + and biomarker - populations, RWD can provide meaningful insight into the prognostic value of these biomarkers that is often requested by HTAs.
5. In working with rare populations, a number of statistical strategies may be used to provide insight into comparative effectiveness. Nonetheless, these strategies will only be acceptable if there is a paradigm shift in reimbursement negotiations, such that the treatment effect could be confirmed with long-term observations in RWD as well as in trial data.

Recommendations for progress in providing access to tumor agnostic treatments

While the focus of the preceding section was on if/how sponsors can generate evidence related to tumor-agnostic therapies, changes are needed at the healthcare system level in order for their efforts to be more successful. We recommend that steps be taken to:

EDUCATE REGULATORS, HTAS, AND PRACTITIONERS

As discussed, sponsors developing drugs in rare and/or tumor agnostic patient populations may need to rely on RWD to generate sufficient evidence to inform regulatory and payer decision-making (Casali, et al., 2015). In light of this, there is a continued need for regulators and HTA bodies to become more familiar with, and accepting of, RWE. (pCODR, 2019; Gemeinsamer Budesausschuss, 2021; Haute Authorité De Santé, n.d.). The concept of a "learning healthcare system" has been proposed as a solution, which will require that data from EHRs and other routinely collected data sources be used to complement evidence based on RCTs (Eichler, et al., 2019).

As the number of drugs targeting specific molecular alterations is growing, NGS will play an increasingly important role in characterizing the diagnosis and in driving treatment regimens towards more personalized approaches that rely on more than tumor histology and anatomical location. However, the use of NGS is currently not universal. Further training and education of medical practitioners is needed.

DEFINE EFFICACY FOR TUMOR-AGNOSTIC DRUGS

It is difficult for HTA bodies to know how a therapy's efficacy compares to that of existing treatments. To our knowledge, there is no guidance from EUnetHTA that outlines the targets for evaluating the value proposition of a tumor-agnostic drug. However, two scales that are available for guiding clinicians and health authorities could be used as models. The first is the ESMO Scale of Clinical Actionability for molecular Targets (ESCAT), and the second is ESMO-MCBS version 1.0 (Cherny et al., 2015)

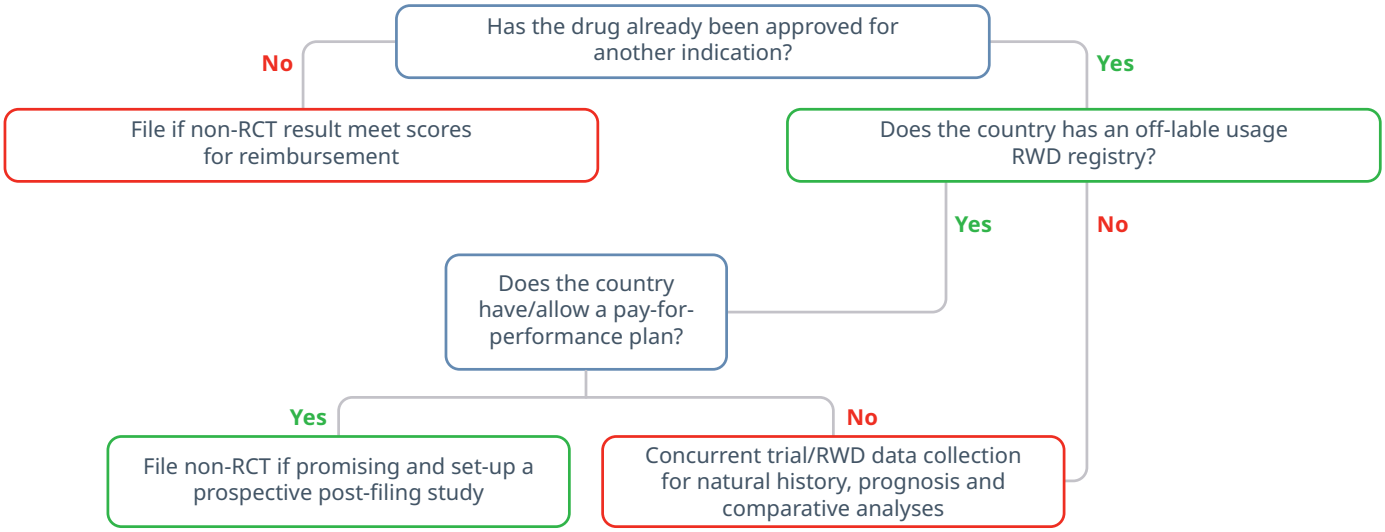
In a setting of a rare and tumor-agnostic population, a head-to-head comparison with an external control is unlikely to generate unequivocal information on the treatment effect. Thus, other comparative methods for showing the validity and consistency of results should be planned as part of generating evidence. One of these methods is intra-patient analysis using Growth Modulation Index (GMI) (Gaultney et al., 2021) or PFS ratio (PFSr) as a proxy measure of the treatment benefit and a novel surrogate endpoint for precision oncology trials (Penel, 2013). Although there are limitations of examining intra-patient efficacy. Beyond these methods, there is also a need to agree on and pre-specify the degree of variability or size of the delta that is reasonable for a tumor agnostic drug. This entails both acknowledging that setting a unique efficacy bar is very difficult and accepting some level of uncertainty and heterogeneity.

CREATE REIMBURSEMENT PATHWAYS AND CASE STUDIES

In many countries, regulatory agencies have made advances in how they assess and interpret evidence associated with tumor agnostic therapies but this has not yet led to positive HTA recommendations. The entrectinib experience has shown us that successful reimbursement schemes depend on the ability to deal with uncertainties through managed-entry agreements (MEAs) rather than on specific points related to the quality of the data itself.

In several countries, the HTA reimbursement pathway does not allow for conditional reimbursement with subsequent evidence development and the ability to address evidence uncertainties with MEAs or outcomes-based payment arrangements (e.g., pay-per-performance). For tumor-agnostic therapies with single-arm trials, there needs to be a more adaptive reimbursement pathway at the country level that evolves with the evidence (including RWD), such as schematized in Figure 1. The Drug Access Protocol (DAP) developed by Dutch oncologists, insurers, and the healthcare public institute is an example of such an innovative pathway, whereby pharmaceutical companies are not paid if the treatment does not work, as determined on the basis of RWD data collection (NL Times, 2021).

Figure 1. Example of possible reimbursement pathways algorithm for better shaping evidence generation for future promising tumor-agnostic therapies Does the country have non-RCT efficacy guidelines for reimbursement?

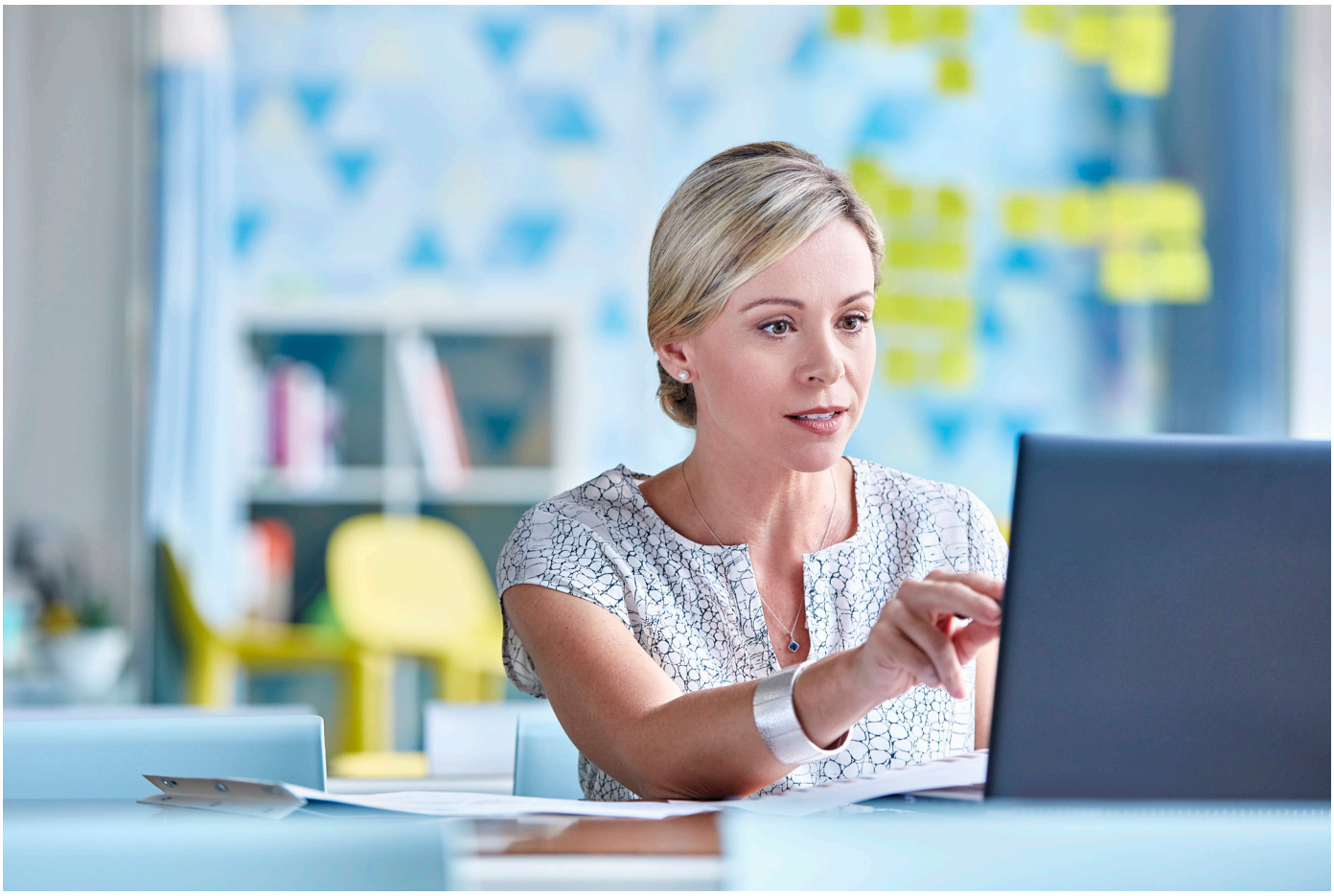


Drugs that are in an advanced development stage for specific indications/tumor-types when they enter a path to be repurposed as tumor-agnostic therapy may utilize different reimbursement pathways and evidence compared to a de novo tumor-agnostic drug.

Oncology societies and health authorities in different countries are developing several initiatives in RWD generation that apply to the tumor-agnostic pathway, including CAPTUR in Canada, DRUP in Netherlands, ProTarget in Denmark, IMPRESS in Norway, MEGALiT in Sweden, and C-CAT in Japan (Radboudumc, 2021). These initiatives focus on evaluating the outcomes of patients treated with targeted therapy in routine care (under off-label indications or limited access schemes agreed upon with pharmaceutical companies) and involve pre-specified prospective data collection to guarantee data quality and credibility. Identifying where/when these initiatives can be repurposed/leveraged in the tumor-agnostic pathway depending on the molecule filing status, still needs to be evaluated and piloted.

ESTABLISH BETTER PARTNERSHIPS

Accomplishing all the above requires better collaboration in the future with payers and other stakeholders (e.g., academia, therapy area experts (TAEs), and other pharmaceutical companies, where possible), to give the opportunity to develop a common framework for evidence generation. This framework should encompass a more common structure across data sources for RWD and EHR and more unified approaches to data collection. Eventually this will result in data sharing platforms that allow partners to leverage commonly created data to bridge evidence gaps, given different perspectives.



Conclusion

There have been challenges and delays in obtaining full approval and reimbursement of tumor-agnostic drugs. Much of the difficulty stems from uncertainties related to the fact that cost-effectiveness must be inferred due to the small number of patients and/or their heterogeneity. Currently, there is a lack of clear guidance outlining what evidence generation framework could reduce these uncertainties. RWD can enrich learnings from single-arm trials to address common uncertainties in natural history, prognostic value, and comparative effectiveness, but there is also a need to understand and account for methodological limitations. In addition to tapping larger and more varied data sources, there is a need to increase data quality and harmonization to obtain meaningful information on a sizable patient population. Doing so will require improved collaboration between drug manufacturers, regulators, and payers to better align on requirements, experiment with new reimbursement pathways, and remove barriers.

References

1. Lozano-Ortega G et al. (2019) PPM11 Tumour-specific randomized controlled trials in rare oncogene-driven cancers: asking for the impossible? Value Health. 01 Nov 19. 22 (3): S838-S839.
2. Pestana, R.C., et al. (2020) Histology-agnostic drug development - considering issues beyond the tissue. Nat Rev. 11 Jun 17(9):555-568.
3. HTx (2021) Next generation health technology assessment (HTx) to support patient-centred, societally oriented, real-time decision-making on access and reimbursement for health technologies throughout Europe.
4. Grigore, B et al. (2020) Surrogate Endpoints in Health Technology Assessment: An International Review of Methodological Guidelines. PharmacoEconomics. 23 Jun. 38: 1055–1070.
5. Casali et al. (2015) Rare Cancers Europe (RCE) methodological recommendations for clinical studies in rare cancers: a European consensus position paper. Ann Oncol. 26(2): 300-306.
6. Hierro, C. et al. (2019) Agnostic-histology approval of new drugs in oncology: are we already there? Clin Cancer Res, 29 Nov. 2019 25 (11): 3210–3219.
7. van Waalwijk S.B. et al. (2019) Personalised reimbursement: A risk-sharing model for biomarker-driven treatment of rare subgroups of cancer patients. Ann Oncol. 30(5): 663-665.
8. EMA (2016). Single-arm trials in cancer drug evaluation. Problem statement EMA workshop on single-arm trials in cancer drug evaluation, EMA, London, 30 June.
9. EUnetHTA (2015). Comparators & comparisons. Criteria for the choice of the most appropriate comparator(s). Summary of current policies and best practice recommendations. Jan 14 2015.
10. Gualtney, J. et al., (2021). Developing a framework for the health technology assessment of histology-independent precision oncology therapies. Appl Health Econ and Health Policy. 19, 625–634 (2021).
11. Smyth, L. M. et al. (2020) Characteristics and outcome of AKT1E17K-mutant breast cancer defined through AACR Project GENIE, a clinicogenomic registry. The AACR Project GENIE Consortium.
12. Shao, C. et al. (2020). Prevalence of high tumor mutational burden and association with survival in patients with less common solid tumors. JAMA Netw Open. 2020 Oct 1;3(10): e2025109.
13. Bazhenova et al (2021). TRK fusion cancer: patient characteristics and survival analysis in the real-world setting. Target Oncol. 2021 May;16(3):389-399.
14. Hyman et al. (2015). Emurafenib in multiple nonmelanoma cancers with BRAF V600 mutations, N Engl J Med. Aug 20;373(8):726-36.
15. Kopetz, et al. (2015). Phase II pilot study of vemurafenib in patients with metastatic BRAF-mutated colorectal cancer, J. Clin. Oncol., 33: 4032-4038.
16. Sung et al. (2020). Tumor mutational burden as a tissue-agnostic biomarker for cancer immunotherapy. Expert Rev Clin Phar. Volume 14, 2021 - Issue 2.
17. Stransky, N. et al. (2014) The landscape of kinase fusions in cancer. Nat Commun 2014; 5:4846.
18. Okamura, R. et al. (2018) Analysis of NTRK Alterations in Pan-Cancer Adult and Pediatric Malignancies: Implications for NTRK-Targeted Therapeutics, JCO Precis Oncol 2, 1-20
19. European Medicines Agency (n.d.) https://www.ema.europa.eu/en/medicines/human/EPAR/rozlytrek
20. European Medicines Agency (n.d.) https://www.ema.europa.eu/en/medicines/human/EPAR/vitrakvi

21. Brogaard, N. et al, (2022) Learnings from the assessments of entrectinib and larotrectinib: health technology assessment challenges associated with tumour-agnostic therapies. OHE Consulting Report, London: Office of Health Economics. Jan 2022.

22. Rosen, E.Y., et al. (2020). TRK fusions are enriched in cancers with uncommon histologies and the absence of canonical driver mutations. Clin Cancer Res. Apr 1;26(7):1624-1632.

23. Demetri, G. et al. (2021). Characteristics and outcomes of patients (pts) with NTRK fusion-positive (NTRK+) metastatic / locally advanced (LA) solid tumours receiving non-TRK inhibitor (TRKi) standard of care (SoC), and prognostic value of NTRK fusions in clinical practice. Ann Oncol (2021) 32 (suppl_5): S382-S406.

24. Bazhenova et al (2021). TRK fusion cancer: patient characteristics and survival analysis in the real-world setting. Target Oncol 2021 May;16(3):389-399.

25. Harder, V.S. et al, (2010) Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. Psychol Methods. 15:234-249.

26. Stuart, E.A. et al, (2013) Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. J Clin Epidemiol. 66(8 Suppl):S84-S90.

27. Davies, J. et al. (2018) Comparative effectiveness from a single-arm trial and real-world data: alectinib versus ceritinib. J Comp Eff Res. Sep;7(9):855-865.

28. Ghadessi, M. et al, (2020) A roadmap to using historical controls in clinical trials – by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG). Orphanet J Rare Dis.15:69.

29. Sammon, C.J. et al. (2020). Real-world evidence and nonrandomized data in health technology assessment: using existing methods to address unmeasured confounding? J Comp Effect Res. 9(14) Aug 1 2020.

30. Thompson, D. (2021). Replication of randomized, controlled trials using real-world data: What could go wrong? Value Health. 2021 Jan;24(1):112-115.

31. Brown, S. (2022). Implications of Selection Bias Due to Delayed Study Entry in Clinical Genomic Studies JAMA Oncol. Feb 1;8(2):287-291.

32. pCODR (2019) Expert review committee final recommendation. https://cadth.ca/sites/default/files/pcodr/Reviews2019/10159LarotrectinibNTRK%2BSolidTumours_fnRec_REDACT_31Oct201_ChairApproved_final.pdf

33. Gemeinsamer Bundesausschuss (2021) https://www.g-ba.de/downloads/40-1465-6479/2020-04-02_AM-RL-XII_Larotrectinib_D-495_TrG_EN.pdf.

34. Haute Autorité De Santé (n.d.) https://www.has-sante.fr/jcms/p_3193768/en/vitrakvi.

35. Eichler et al (2019). Data rich, information poor: can we use electronic health records to create a learning healthcare system for pharmaceuticals? Clin Pharmacol Ther. Apr 2019 105(4).

36. ESMO (n.d.) https://www.esmo.org/policy/esmo-scale-for-clinical-actionability-of-molecular-targets-escat.

37. ESMO (n.d.) https://www.esmo.org/guidelines/esmo-mcbs/scale-evaluation-forms-v1.0-v1.1#:~:text=In%202015%20ESMO%20launched%20the,to%20high%20value%20cancer%20treatments.

38. Cherny, N. et al. (2015). A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). Ann Oncol 26: 1547–1573.

39. Penel, N. et al, (2013) Growth modulation index as metric of clinical benefit assessment among advanced soft tissue sarcoma patients receiving trabectedin as a salvage therapy. Ann Oncol. 2013 Feb;24(2):537-542.

40. NL Times (2021). New cancer medicine more quickly available to Dutch patients. Jan 19 2021. https://nltimes.nl/2021/01/19/new-cancer-medicine-quickly-available-dutch-patients.

41. Radboudumc (2021). Dutch-Nordic Alliance for Precision Cancer Medicine. Nov 2021. https://www.radboudumc.nl/en/news-items/2021/dutch-nordic-alliance-for-precision-cancer-medicine-launched

Table Appendix 1. Snapshot of natural history and NTRK+ prognosis results with the Flatiron clinico-genomic dataset (data cut-off December 2020, follow-up June 2021) with insights for evidence generation. https://oncologypro.esmo.org/meeting-resources/esmo-congress/characteristics-and-outcomes-of-patients-pts-with-ntrk-fusion-positive-ntrk-metastatic-locally-advanced-la-solid-tumours-receiving-non-trk

	RESULTS	INSIGHTS FOR EVIDENCE GENERATION
SAMPLE SIZE	N=36 with inclusion/exclusion criteria; -8 patients in/our network with data not usable (N=28). (includes 8 larotrectinib and 2 entrectinib treated patients observed and removed from dataset)	• US centric • With current small patient numbers, conclusive findings are not possible for full understanding of the natural history of patients, and further data and additional sources are needed to capture well the heterogeneity of NTRK+ fusion patients • Dataset is unlikely to grow considerably as it is expected that all NTRK+ will be treated with either larotrectinib or entrectinib in future in the US
TUMOR TYPES	A total of 10 different tumor types spread across our 28 patients captured. Most frequent tumors were colorectal (32%), sarcoma (21%) and NSCLC (18%). Other tumors were only represented by one to two patients (cancer of primary unknown [CUP], endometrial, stomach, billary, uterus, salivary glands)	Clinical practices, over-representation common cancers
NGS TESTING	Mean time from advanced metastatic to reported FMI-NGS results was 151.21 days (SD 245.20).	Lung, sarcoma, and CRC seems now routine to be tested at advanced diagnoses for treatment decision
TREATMENTS	Of those with available information, 71% received up to two lines of antineoplastic treatments since initial diagnosis until FMI-NGS report date	Chemotherapy in combination with non-TRKi or immunotherapy in line or as separate line and could be considered as comparator under SoC; most patients have received either 1 or 2 lines therapy previous NGS test.
TIME PERIOD	NGS test frequency seems raising from <10% in 2014 to 21% in 2019	In retrospective data, patients may not have "contemporary" treatment
OS	Median OS in the NTRK+fusion cancer patients is 10.2 (months) with 95%CI confidence intervals 7.2-14.1 months	Could inform value proposition for overall efficacy acceptability for example: HR=0.7, i.e., for 3 months added OS if OS SoC is under 7 months HR=0.8, i.e., for 3 months added OS if OS SoC is 14 months -> for HR=0.5 at least 5 months for 10 months SoC
HR NTRK- VS NTRK-	HR: 1.6 (1.0-2.5) Matching scheme: tumor type direct; Propensity score near neighbor on: age, smoking status, practice type, number of lines of antineoplastic treatments since initial diagnosis therapy until FMI-NGS report date, stage at diagnosis, reported time between advanced/metastatic and reported test., co-mutations, MSI-H, TMB-H	• Small negative prognostic value; `but provides indirect evidence that if targetable drug shows benefit in single-arm trial likely due to an effect of the drug (no confounding by a third variable) • It was possible to leverage a large biomarker negative population from the same source (15000 NTRK- fusion patients) and give insight on important prognostic/ confounders to retain.

Table Appendix 2. Key design elements of pilot study of head-to-head comparison of clinical trial with RWD as external control with insights for evidence generation

ITEM	DEFINITION	INSIGHTS FOR EVIDENCE GENERATION
OBJECTIVE	To explore relative efficacy of entrectinib in NTRK fusion-positive patients independent of underlying histology	There is a known efficacy-effectiveness gap from trial and routine care that may or may not be controlled with matching by prognostic factors or confounders
HYPOTHESIS	There is a clinical meaningful difference in treatment effect between entrectinib and SoC across all tumor and lines of treatment;	The clinically meaningful difference can only be defined as a Hazard Ratio on the point estimate equal or below a certain level (i.e., clinical benefit of ESMO scale,), across lines and tumors
SOURCES AND OVERALL STUDY DESIGN	Entrectinib trial arm (STARTRK-2) and FH-CGDB as external control; Comparison of clinical trial vs real-world outcomes	Only one RWD source so far available; USA centric
POPULATIONS	Patients with advanced NTRK fusion positive tumors, any tumor, any line	FH-CGDB can only capture FMI tested patient, avoid misclassification fusion compared to trial but lack representativeness of patient tested with another test
TREATMENTS	Intervention: entrectinib External Control: real-world standard of care treatments	Standard of care varies by tumor and line; how representative the SoC captured in one dataset is of all future treated patients is unclear
SAMPLE SIZE	Initial population: 57 entrectinib patients; 27 Flatiron-CGDB patients Compared population: 56 entrectinib trial patients vs 12 Flatiron-CGDB matched patients (only includes trial patients with tumors observed in Flatiron-CGDB)	Assuming a 1:3 case-control match is minimally necessary for meaningful comparison, there is currently no availability of non-trial patients to conduct such comparison
ENDPOINTS	Overall Survival	OS is relevant for HTA bodies; confirmation with other outcomes will add to robustness results; need additional validity studies on real-world PFS and real-world response
INDEX TIME (S)	Entrectinib trial: Time from start of entrectinib; Flatiron-CGDB time from NGS report	Bias added due to difficulty match index times; conservative estimate
STATISTICAL CONSIDERATIONS	Matching by tumor type and by near neighbor propensity score matching; The PS was estimated using logistic regression on the basis of a minimum set of four a priori selected prognostic variables categorized that included age (>=65), time from initial diagnosis to index (>6 months), stage at initial diagnoses (>=III/IV), and number of prior lines of therapy (>=2) (after advanced/metastatic diagnoses to index date); sensitivity on index time for Flatiron-CGDB	Not all variables that may be considered prognostic can be included, although more important ones are included and can act as proxy; ECOG is highly missing (50%)
ASSUMPTIONS	• Overall efficacy only; evidence for supporting efficacy claim given by a point estimate of the HR<1 • Trial and external sources come from a same underlying population of patients • Underlying model assumptions are met (i.e., proportional hazard model)	• Relaxing of assumption is needed as we know upfront some are not valid • Hypothesis may not be testable in the formal statistical sense (Type 1 error)

CONTACT US

iqvia.com