NIPH
Norwegian Institute of Public Health

REPORT

2022

HEALTH TECHNOLOGY ASSESSMENT:

Tests for the detection of NTRK gene fusions in patients with locally advanced or metastatic solid tumours

Publisher Norwegian Institute of Public Health, Division for Health Services

Title Tests for the detection of NTRK gene fusions in patients with locally advanced or metastatic solid tumours. A Health Technology Assessment.

Norwegian title Tester for deteksjon av NTRK genfusjoner hos pasienter med lokalavanserte eller metastatiske solide svulster. En metodevurdering.

Authors Vida Hamidi, p*roject leader, senior researcher, Norwegian Institute of Public Health*
Gerd M. Flodgren, *senior researcher, Institute of Public Health*
Jose F. Meneses-Echavez, *researcher, Norwegian Institute of Public Health*
Julia Bidonde*, senior researcher, Norwegian Institute of Public Health*

ISBN 978-82-8406-268-6

Project number ID2019_119 and ID2019_029

Type of report Health Technology Assessment

No. of pages 73 (104 including appendices)

Client Commissioning Forum for Nye metoder

Subject heading (MeSH) Neutrotrophic Tyrosine Receptor Kinase, NTRK, NTRK-fusion, solid tumours, pharmacogenomic testing, IHC, FISH, NGS, RT-PCR, sensitivity, specificity, concordance, accuracy, costs and cost analysis

Citation Hamidi V, Flodgren GM, Meneses-Echavez JF, Bidonde J. Tests for the detection of NTRK gene fusions in patients with locally advanced or metastatic solid tumours. A Health Technology Assessment. Norwegian Institute of Public Health, 2022.

Key messages

The Norwegian Institute of Public Health has been commissioned to assess molecular tests for the identification of NTRK gene fusions in locally advanced or metastatic solid tumours. Less than 1% of solid tumours have somatic NTRK gene fusions, with higher prevalence in younger children than in adults. Accurate and reliable detection of NTRK fusions is important for identification of people who may benefit from drug treatment (e.g., entrectinib and larotrectinib), as well as NTRK fusion negative patients, to avoid provision of unnecessary and costly medications.

We included nine studies comparing one or more analytical techniques (IHC, FISH, RT-PCR, NGS) for the detection of NTRK gene fusions that reported test accuracy data. Five narrative reviews, and two expert opinion papers provided feasibility data. Experts were contacted for cost information. The results of this HTA show that:

- Test accuracy was mostly inadequate, and reporting was poor.
- Un-pooled results including six test comparison suggest varying test accuracy mostly for single gene testing (e.g., IHC), across different types of solid tumours and NTRK fusions.
- The results suggest higher sensitivity of RNA-NGS than DNA-NGS in detecting NTRK gene fusions, especially for fusions with large intronic regions (NTRK2, and NTRK3).
- While there are advantages and limitations for all the tests, single gene testing may be unfeasible, especially when the number of actionable biomarkers relevant for testing are increasing.
- Due to a tendency for false positive staining NTRK fusions positive with IHC needs confirmation with other molecular methods (e.g., RT-PCR or RNA-NGS).
- The development of a testing algorithm for the detection of NTRK fusions depends on accessibility of testing modalities, economic considerations, histology and turnaround time.
- The cost associated with NGS testing will significantly decrease when parallel tests are performed for several biomarkers (using gene panels) from multiple patients. At present, the capital and infrastructure as well as maintenance costs are higher for NGS than the other diagnostic methods.
- NTRK fusions can be detected in many different types of advanced solid tumours, we estimated that between 10,000-11,100 people may be eligible for NTRK testing in Norway each year. The cost for testing common solid tumours with a low frequency of NTRK fusions using IHC as pre-test with NGS confirmation were estimated to be 16.1-18.0 million Norwegian kroner (NOK). The costs for testing rare tumours with a high frequency of NTRK fusions with NGS were estimated to be about NOK 1.2 million.
- Future research should focus on conducting larger cohort studies with well-defined patient populations, that follows the patients from testing (or no testing), through treatment and final outcomes. Further, robust and replicable methods, as well as a reporting standard checklist, should be used for increased clarity.

Title:
Tests for the detection of NTRK gene fusions in patients with locally advanced or metastatic solid tumours

---

Type of publication:
Rapid Health Technology Assessment

Publisher:
Norwegian Institute of Public Health

Updated:
Last search for literature May 2021

Clinical experts:
Emilius AM Janssen, SUS
Hege LG Russnes, OUS
Lars Helgeland, UiB
Tormod K Guren, OUS
Åslaug Helland, OUS

Hovedbudskap

Folkehelseinstituttet har på oppdrag fra Bestillerforum for nye metoder evaluert molekylære tester for identifisering av NTRK genfusjoner hos pasienter med lokalavanserte eller metastatiske solide svulster. Mindre enn 1 % av alle solide svulster har somatiske NTRK genfusjoner, med høyere prevalens blant yngre barn enn blant voksne. Tester som identifiserer hvilke pasienter som kan ha nytte av målrettet medikamentell behandling er viktige for adekvate behandlingsbeslutninger.

Vi inkluderte ni studier som sammenlignet én eller flere analytiske metoder (IHC, FISH, RT-PCR, NGS) for deteksjon av NTRK genfusjoner og rapporterte data om testnøyaktighet. Fem narrative oversikter og to ekspertuttalelser ga data om anvendbarhet. Eksperter ble kontaktet for kostnadsinformasjon. Resultatene av denne metodevurderingen viser at:

- De vurderte testenes nøyaktighet var stort sett utilstrekkelig dokumentert, og resultatrapporteringen i de inkluderte studiene var ikke tilfredsstillende.
- Resultater fra enkeltstående studier av seks ulike sammenligninger viste varierende testnøyaktighet på tvers av ulike solide svulster og NTRK-fusjoner.
- RNA-NGS har muligens bedre sensitivitet enn DNA-NGS for påvisning av NTRK genfusjoner, spesielt fusjoner med store introniske regioner (NTRK2 og NTRK3).
- Alle testmodaliteter har sine fordeler og begrensninger, men enkeltgentesting vil sjelden være hensiktsmessig når antall biomarkører som skal testes øker.
- Positive NTRK-fusjoner påvist med IHC må bekreftes med andre molekylære metoder (f.eks. RT-PCR eller RNA-NGS).
- Utviklingen av en testalgoritme for påvisning av NTRK-fusjoner avhenger av tilgjengeligheten til testmodaliteter, økonomiske hensyn, histologi og tidsbruk.
- Kostnader knyttet til NGS-testing reduseres betydelig hvis man kan utføre parallelle tester av flere biomarkører fra flere pasienter, men per i dag er kapital-, infrastruktur og vedlikeholdskostnadene høyere for NGS enn andre diagnostiske metodene.
- NTRK-fusjoner kan påvises i ulike typer avanserte solide svulster. Mellom 10 000-11 100 personer kan anslagsvis kvalifisere for NTRK-testing i Norge hvert år. Kostnaden for å teste vanlige forekommende solide svulster med lavfrekvente NTRK-fusjoner ved bruk av IHC som pretest og bekreftelse med NGS estimeres til 16,1-18,0 millioner norske kroner (NOK). For testing av sjeldne svulster med høyfrekvente NTRK-fusjoner ved bruk av NGS anslås kostnaden til ca. 1,2 millioner kroner.
- I fremtidig forskning bør man prioritere gjennomføring av store kohortstudier med veldefinerte pasientpopulasjoner der man følger pasientene fra testing (eller ingen testing), gjennom behandling og til sluttresultater. Man bør ta i bruk mer robust og reproduserbar metodikk samt standardiserte rapporteringsmaler med mål om å tydeliggjøre resultatene og bedre kvaliteten til dokumentasjonen.

Tittel:
Tester for deteksjon av NTRK genfusjoner hos pasienter med lokalavanserte eller metastatiske solide svulster

---

Publikasjonstype:
Forenklet metodevurdering

Hvem står bak denne publikasjoner:
Folkehelseinstituttet på oppdrag fra Bestillerforum for nye metoder

Litteratursøk avsluttet:
Mai 2021

Kliniske fageksperter:
Emilius AM Janssen, SUS
Hege LG Russnes, OUS
Lars Helgeland, UiB
Tormod K Guren, OUS
Åslaug Helland, OUS

Table of contents

Preface

The Commissioning Forum ("Bestillerforum for nye metoder" in Norwegian), representing the four Regional Health Authorities (RHFs) through the National System for Managed Introduction of New Health Technologies within the Specialist Health Service in Norway (Nye Metoder), has commissioned the Norwegian Institute of Public Health (NIPH, Folkehelseinstituttet) to conduct an assessment of relevant diagnostic tests for identification of neurotrophic tyrosine receptor kinase (NTRK) gene fusions patients with locally advanced or metastatic solid tumours. This assessment is conducted based on two separate commissions, ID2019_119 and ID2019_029, where NIPH has responsibility for assessing the diagnostic tests used to identify NTRK gene fusions and the Norwegian Medicine Agency (NoMA) has responsibility for conducting two single technology assessments of the relevant treatments, entrectinib and larotrectinib.

This health technology assessment (HTA) includes a summary of original papers and reviews reporting on the sensitivity, specificity, concordance, feasibility, and cost analysis of four different tests (ImmunoHistoChemistry (IHC), Fluorescence in situ hybridization (FISH), Reverse transcription-polymerase chain reaction (RT-PCR) and Next-Generation Sequencing (NGS)) for the identification of NTRK gene fusions in patients with locally advanced or metastatic solid tumours. This report aims to support well-informed decisions in health care that lead to improved quality of services.

The internal project group included the following members affiliated with the Norwegian institute of Public Health:

- Vida Hamidi (VH), Senior researcher (health economist), project leader
- Gerd M Flodgren (GMF), Senior researcher, lead systematic review
- Jose Francisco Meneses Echavez (JFME), Researcher, systematic review
- Julia Bidonde (JB), Researcher, systematic review
- Elisabet Hafstad (EH), Research librarian, literature search

We would like to thank our external clinical experts Emilius AM Janssen (Professor, group lead, Stavanger University Hospital), Hege Russnes (Senior consultant/group lead, Oslo University Hospital), Lars Helgeland (Associate professor, Haukeland University Hospital), Tormod K Guren (Oncologist, Oslo University Hospital) and Åslaug Helland (Professor, senior oncologist, Oslo University Hospital) for their

expertise in this project. We also wish to acknowledge Øyvind Melien for his contribution to this project.

All the authors of this HTA and the clinical experts declared no conflict of interest.

We will emphasise that although the clinical experts have contributed with valuable input and comments, NIPH is solely responsible for the content of this report.

Kåre Birger Hagen
Director of Reviews and Health Technology Assessments

Kjetil Gundro Brurberg
Department director of Reviews and Health Technology Assessments

Vida Hamidi
Senior Researcher and Project leader

Abbreviations

ALK	Anaplastic Lymphoma Kinase fusion oncogene
AUROC	Area Under the Receiver Operating Characteristic Curve
CI	Confidence Interval
CRC	Colorectal carcinoma
DRG	Diagnosis-related group
EGAPP	Evaluation of genomic applications in practice and prevention
ELSI	Ethical, legal, and social implications
EMA	European Medicine Agency
ESMO	European Society for Medical Oncology
EUnetHTA	European Network for Health Technology Assessment
FDA	American Food and Drug Administration
FF	Fresh Frozen
F1CDx	The FoundationOne CDx assay
FFPD	Formalin-fixed Paraffin Embedded samples
FISH	Fluorescence in Situ Hybridisation
HTA	Health Technology Assessment
IHC	Immuno-Histo-Chemistry
INAHTA	International Network of Agencies for HTA
NGS	Next Generation Sequencing
NIPH	Norwegian Institute of Public Health
NordiQc	Nordic Immunohistochemical Quality Control
NoMA	Norwegian Medicine's Agency
NSCLC	Non-small cell lung carcinoma
NTRK	Neurotrophic tyrosine receptor kinase
Pan Trk	Pan-tropomyosin-related-kinase
PTC	Papillary Thyroid Cancer
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RET	REarrangement during Transfection (RET) oncogene
RHF	Regionale Helseforetak (Norwegian)

ROS1 gene	Proto-oncogene tyrosine-protein kinase fusion protein
RTK	Receptor tyrosine kinase
RT-PCR	Reverse transcription polymerase chain reaction
PD-LI	Programmed Death Ligand 1
PICO	Population, Intervention, Comparison, Outcomes
SC	Secretory carcinoma
SR	Systematic review
TAT	Turn-around time
TKI	Tyrosine kinase inhibitor
TrkA	Tropomyosin receptor kinase A

Objectives

The main objective of this assessment was to summarise available evidence on the analytical validity, the clinical validity, and the clinical utility of relevant diagnostic tests, including both single gene (IHC, FISH, RT-PCR) and multigene biomarker analyses (NGS) for the detection of neurotrophic tyrosine receptor kinase (NTRK) gene fusions in patients with locally advanced or metastatic solid tumours. More precisely we aimed to answer the following research questions:

- How accurately and reliably do each of these tests detect the biomarker in the laboratory (technical performance)?
- How accurately and reliably do each of these tests detect the biomarker in samples from patients (e.g., tumour tissue, circulating cells, or cytology samples) with different types of solid tumours?
- How well do each of these tests predict the effectiveness of treatment (e.g., shrinking of the tumour, or slowing down the disease process)?
- How well do each of these tests predict outcomes of importance to the patient (e.g., overall survival, and quality of life)?
- What are the potential adverse effects of using these tests to guide treatment decisions affecting patients?
- What are the advantages and limitations of the different tests (i.e., the feasibility of tests in terms of biological tissue requirements, turnarond time, invasiveness, training/expertise needed for running the analyses or interpreting the test results)

We have also assessed the cost related to the use of these diagnostic methods.

Service delivery and organisational aspects related to test services in Norway, the ethical, legal, and social implications (ELSI), and patient preferences related to pharmacogenomic testing have been addressed in a previous publication from the Norwegian Institute of Public Health (1).

Background

General

Precision medicine (PM) is a term that is increasingly being used to describe treatments, including therapeutic agents, tailored to individual patients or groups of patients (2). The overall goal is to match pharmacological therapies to individuals to ensure that they receive effective treatment with minimal toxicity. This is particularly important for cancer patients who may have a limited life expectancy (3).

The most significant aspect of a PM approach, within the field of oncology, involves the identification of a 'biomarker' associated with a particular cancer type. A biomarker is a unique mutated nucleic acid sequence, protein, glycoprotein, or group of proteins, expressed by the tumour cells but not normally by healthy cells (2). There are four main types of biomarkers: pre-disposition (indicating the likelihood of developing the disease), diagnostic (used to confirm the patient has a particular cancer), prognostic (suggesting how cancer may develop in the individual), and predictive (determining which cohort of patients may benefit from a particular drug therapy) (2).

The potentially improved patient outcomes provided by PM, depend on the accurate identification of patients for treatment, based on the predictive biomarker testing (4). Hence, there is a natural dependency that exists between biomarker-based treatment and test. Unfortunately, designing a validated diagnostic assay to identify the right patients for treatment does not guarantee accurate detection of the right patient population and subsequent delivery of treatment (5). Effective use of biomarker tests and applying high-quality testing standards are fundamental to deliver precision medicine.

There has been a steady growth in the number of genomic tests available for use in healthcare services during the last two decades (6;7). The task of determining the appropriateness of the plethora of different tests is a challenge for both clinicians and policy- and decision-makers (6). The clinical use of reliable tests to guide therapy selection depends on many related processes (i.e., analytical validation, clinical validation, specimen handling, reproducibility, information technology, infrastructure), which can affect the accuracy and reliability of test results and patient safety (8).

Neurotrophic tyrosine receptor kinase (NTRK) gene fusions are an actionable biomarker for cancer therapy and can be found in over 25 different types of cancer, regardless of where they are located in the body (9;10). Two medicines, entrectinib and larotrectinib, were recently approved for the treatment of NTRK fusion by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA). NTRK fusion assessment is therefore expected to become a standard part of management for patients with locally advanced or metastatic solid tumours. Unlike somatic assessment, the detection of NTRK fusions is not straightforward and various test methodologies are proposed for the detection of NTRK fusions (11).

Condition/disease

Epidemiology

There are more than 100 different types of cancer. Hematologic (blood) cancers and solid tumour cancers are two main categories of cancer (12). Solid tumours are abnormal localised masses of tissue that usually does not contain cyst or liquid areas. They can be benign (not cancerous), or malignant (cancerous). Different types of solid tumours are classified according to the type of cells that form them (13). The two major types of cancerous solid tumours are sarcomas and carcinomas. Sarcomas are developed from cells of muscles, bone or fat tissue and carcinomas start from the epithelial cells in the skin or tissues that line or cover internal organs (NICE). Advanced solid tumours can be locally advanced (tumour that has spread to surrounding tissues or lymph nodes but has not yet spread to other parts of the body) or metastatic (tumour that has spread to other parts of the body).

NTRK gene family contains three members, NTRK1, NTRK2, and NTRK3, which produce tropomyosin receptor kinase (TRK) proteins TRKA, TRKB, and TRKC, respectively (9). The TRK proteins are exclusively expressed in human neuronal and extraneuronal tissue where they regulate pain, proprioception, appetite, and memory (9;10;14). Oncogenic gene fusions occur by chromosomal rearrangements of NTRK1, NTRK2, and NTRK3 genes. These gene fusions cause tissue-agnostic overexpression of TRK proteins that affect downstream signalling, which can lead to the uncontrolled growth of cancer cells (9;10).

Less than 1% of solid tumours in children and adults have somatic chromosomal gene fusions involving NTRK genes (15). NTRK fusion-positive tumours prevalence varies by age and cancer type; it has been reported 0.28% in adults (aged ≥18 years) and 1.34% in children (aged <18 years). Prevalence increases with decreasing age, with children <5 years demonstrating the highest incidence); largely as a result of NTRK fusion-positive soft tissue fibrosarcoma (not found in other age groups) (16).

NTRK gene fusions are rare (under 5% frequency) in more common solid tumours (e.g., colorectal (0.7-1.5%), breast, melanoma (0.3%) and lung cancers (0.2%-0.3%)),

but have been detected at high frequencies (over 80%) in some rare cancers (e.g., secretory breast carcinoma, secretory salivary gland cancer, also known as mammary analogue secretory carcinoma of the salivary gland, and congenital mesoblastic nephroma) and in some paediatric cancers (over 90%) (e.g., infantile fibrosarcoma: 91%–100) (17) (Table 1).

Table1. *Frequency and type of NTRK gene fusion cancers in adult and pediatric patients*

	Adult cancers	Paediatric cancers
High frequency (>80%)	• Mammary analogue secretory carcinoma (secretory salivary gland cancer) (NTRK3) • Secretory breast carcinoma (NTRK3)	• Secretory breast carcinoma (NTRK3) • Infantile fibrosarcoma and other mesenchymal tumours (NTRK1,3) • Cellular and mixed congenital mesoblastic nephroma (NTRK1,3)
Intermediate frequency (5%-25%)	• Papillary thyroid cancer (NTRK1,3)	• Papillary thyroid cancer (NTRK1,3) • Spitz tumours (NTRK1,3) • Paediatric high-grade gliomas (NTRK1,2,3)
Low frequency (<5%)	• Appendiceal cancer (NTRK3) • Glioma/glioblastoma (NTRK1,2,3) • Astrocytoma (NTRK2) • Gastrointestinal stromal tumour (NTRK3) • Head and neck cancer (NTRK 2,3) • Lung cancer (NTRK 1,2) • Sarcoma (NTRK1, 3) • Breast cancer (NTRK1,3) • Acute lymphoblastic leukaemia, acute myeloid leukaemia, histiocytosis, multiple myeloma, dendritic cell neoplasms (NTRK3) • Uterine sarcoma (NTRK1,3) • Cholangiocarcinoma (NTRK1) • Pancreatic cancer (NTRK1) • Melanoma (NTRK1,2,3) • Colorectal cancer (NTRK1,3)	• Ganglioglioma (NTRK2) • Astrocytoma (NTRK2)

Sources: Marchio 2019 and Penault-Llorca 2019 (18;19)

TRK fusion proteins are often mutually exclusive of other known fusion proteins involving kinases. Specific NTRK gene fusions are associated with certain tumours, for example, the ETV6-NTRK3 gene fusion is exhibited by 90%–100% of mammary analogue secretory carcinomas and of the secretory breast cancers and is present in most cases of infantile fibrosarcoma and congenital mesoblastic nephroma (19). In contrast, some cancers have many different fusion partners. For example, in lung cancer, seven different gene fusions involving the NTRK1 gene leading to constitutive TRKA tyrosine kinase domain activation have been described (18;19). For more information, see Appendix 1.

Overview of existing treatments

Tumours derived from an NTRK gene fusion are commonly referred to as "TRK fusion cancers" (17). Based on their putative role in cancer cell proliferation, TRK fusion proteins are an active area of investigation and are the molecular target of some approved drugs, including larotrectinib and entrectinib.

Larotrectinib was approved by the FDA in 2018 for treatment of adult and paediatric patients with solid tumours that have a NTRK gene fusion without a known acquired resistance mutation, that are either metastatic or where surgical resection is likely to result in severe morbidity, and who have no satisfactory alternative treatments or whose cancer has progressed following treatment (20). Larotrectinib has also been given *conditional* marketing authorization by the EMA in 2019 (21). The EMA's approval was based upon pooled data from 102 patients across three Phase I and II trials (21).

Entrectinib was approved by the FDA in 2019 for the treatment of adult and paediatric patients 12 years of age and older with solid tumours that have a NTRK gene fusion without a known acquired resistance mutation, are metastatic, or where surgical resection is likely to result in severe morbidity, and have progressed following treatment or have no satisfactory alternative therapy (22). Entrectinib has also received *conditional* marketing authorization from the EMA in 2020 (23). Approval was based on a pooled analysis comprising 93 patients with TRK fusion-positive enrolled across three open-label single-arm phase 1/2 studies (24).

EMA considered treatment of advanced solid tumours with NTRK gene fusions with larotrectinib or entrectinib is of benefit when other treatment is not available or does not work. However, more information is needed on the medicines' effect on tumours in different sites and also when other gene abnormalities are present (24).

Larotrectinib and entrectinib have received marketing authorization in Norway but they are not yet approved by the Decision Forum of the National System for Managed Introduction of New Health Technologies within the Specialist Health Services.[1] Hence, there is at present no approved treatment specifically aimed at patients with NTRK fusion cancer in Norway. The current treatment for solid tumours is based om where in the body the cancer starts and generally includes surgery, chemotherapy, radiotherapy, hormone therapy, immunotherapy, and/or targeted drug therapy (25).

[1] The Norwegian Medicine Agency was commissioned by the Commissioning Forum in the National System for Managed Introduction of New Health Technologies within the Specialist Health Service to perform two single technology assessments of larotrectinib and entrectinib for treatment of patients with NTRK gene fusion positive locally advanced or metastatic solid tumours.

Molecular tests for detection of NTRK gene fusions

For optimal clinical efficacy of TRK inhibitors, an effective diagnostic strategy to detect NTRK gene fusions in tumour samples is essential to guide treatment selection (19). Since, there are three different NTRK genes, a variety of potential fusion partners and a few possible breakpoints at which different exons of the NTRK tyrosine kinase would join the fusion partner, screening for NTRK fusions is complex (11). The choice of the test for the detection of NTRK gene fusions will depend on the frequency and type of the NTRK gene fusion in a particular tumour, as well as time-consuming, material-dependent, costs and availability of each of the test.

Methods that may be used to directly or indirectly detect the presence of a gene fusion in tumour tissue samples include immunohistochemistry (IHC), fluorescence in situ hybridisation (FISH), reverse transcriptase-polymerase chain reaction (RT-PCR), and next-generation sequencing (NGS) using deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) (18).

While IHC testing is sometimes used, the new guidelines recommended that confirmatory testing for NTRK gene fusions should be performed at the molecular level (DNA-based NGS or RNA-based NGS, FISH, and RT-PCR) (18). Historically, gene fusions have been assayed by FISH and RT-PCR, and FISH assays for the detection of the ETV6–NTRK3 fusion gene are commercially available. However, given the multitude of 5´ partners involved in NTRK1/2/3 fusion genes, assays that allow for the detection of multiple variants in a single test, including DNA-based NGS or RNA-based NGS approaches, have been widely used in large academic centres in North America and European countries. At the same time, the adoption of these NGS-based methods in other contexts has proven challenging, given the costs for the implementation and running of the assay, limited reimbursement by the public, need for bioinformatics expertise, and relatively longer turnaround time (1–3 weeks) (18). As presented, each type diagnostic method has its own distinct advantages and limitations however, to our knowledge, there is still a lack of systematic review to assess the accuracy of different tests for the detection of NTRK gene fusions.

In 2019, the European Society for Medical Oncology (ESMO) Translational Research (TR), in collaboration with Precision Medicine Working Group (PM WG) reviewed the available methods for the detection of NTRK gene fusions. A consensus on the most reasonable strategy to adopt when screening for NTRK fusions in oncologic patients was sought, and further reviewed and approved by the ESMO TR and PM WG and the ESMO leadership. The recommended testing algorithm by ESMO is based on the histology-based triage (18;26). In this report, a brief review of the proposed algorithms for NTRK gene fusion testing to identify patients who may benefit from therapies targeting TRK fusion proteins, published in the relevant literature including ESMO recommendation was presented.

Currently, NTRK testing is not routinely done for all solid tumours in Norway. However, NGS is used at most Norwegian university hospitals to detect NTRK gene fusions (personal communication).

According to a survey performed in 2020, a majority of the Norwegian hospitals have invested in NGS technology, and it is expected that NGS will be available at all hospitals in a short time (27). The survey reported on the relevant challenges with implementing NGS diagnostics in cancer including lack of personnel, small area, lack of guidelines on which genes to analyse (size of gene panel), and which findings to report. Other challenges with NGS diagnostics were related to analysis were poor quality DNA and RNA, due to the type of samples most often used (i.e., formalin-fixed, paraffin-embedded, FFPE). The survey also showed that panel size and reporting of results varied across the hospitals.

Companion diagnostics

The FDA has approved the Foundationone®CDx assay (F1CDx) (28) to be used as a companion diagnostic to identify fusions in NTRK genes, NTRK1, NTRK2 and NTRK3, in DNA isolated from tumour tissue specimens from patients with solid tumours eligible for treatment with larotrectinib (28).

F1CDx is a next-generation sequencing-based in vitro diagnostic device that is capable of detecting several mutations in addition to NTRK gene fusions in 324 genes and select gene rearrangements (28;29). F1CDx does not have coverage of NTRK3 intronic regions, while the most common rearrangement gene partner of NTRK3, which is ETV6, is covered by F1CDx (30).

The supplier of entrectinib, Roche, has submitted F1CDx to the FDA for approval as a companion diagnostic for entrectinib however an FDA-approved companion diagnostic for entrectinib is not available at this time (31).

Why is it important to conduct this assessment?

In this HTA, we have summarised the evidence of the accuracy of different tests relevant for the detection of NTRK gene fusions and described the advantages and limitations of these tests. In addition, we have conducted an economic evaluation and estimated the costs associated with these diagnostic methods in Norway. Regarding information on organization and delivery of services, and patient preferences related to molecular testing, we have relied on the results of a recent publication from NIPH (1). This assessment was conducted to assist decision-makers in making informed decisions regarding the delivery and organisation of molecular tests services in Norway.

Methods

A systematic review of the literature was conducted in response to this commission. The purpose of the molecular tests under study was to predict the treatment response or adverse events. We used a combination of the EGAPP framework (32;33), and the extended framework described by Pitini et al. to guide our assessment (34). A glossary is provided in Appendix 2.

Literature search

Research librarian Elisabet Hafstad (EH) developed the search strategy with input from the authors, planned and ran the electronic searches in the following data bases in April 2020 and May 2021:

- Cochrane Central Register of Controlled Trials (Wiley) - April 2020
- Embase (Ovid) – April 2020, May 2021
- MEDLINE (Ovid) – April 2020, May 2021
- ClinicalTrials.gov (US National Institutes of Health) – April 2020, May 2021
- International Clinical Trials Registry Platform (WHO) – May 2021 (not searched in April 2020 due to technical problems)
- PROSPERO International Prospective Register of Systematic Reviews (National Institute for Health Research, UK) - April 2020
- EUnetHTA POP database (EUnetHTA) - April 2020

The literature does not recommend using method filters for study design in searches for studies on diagnostic tests (35), and our search strategy, therefore, consisted only of words and variants for NTRK fusion searched in the text (title and summary) and if available, in the controlled vocabulary. The search was not limited by language or year of publication. We excluded articles describing animal research. In the update search, May 2021, we also did not collect conference abstracts. The complete search strategy is provided in Appendix 3.

Inclusion criteria

We used the PICO (population, intervention, comparison and outcomes) framework to describe the inclusion criteria (36).

Table 2. *PICO-criteria for inclusion of studies*

PICO	
Population:	Adults and children with any type of locally advanced or metastatic solid tumours
Intervention (index test(s)):	-Immunohistochemistry (IHC), -Fluorescence In Situ Hybridization (FISH), -Reverse Transcription Polymerase Chain Reaction (RT-PCR), -DNA- and RNA based Next Generation Sequencing (NGS)
Comparison (reference test(s)):	Head-to-head comparisons of the tests listed above
Outcomes:	Analytical validity, clinical validity, clinical utility, feasibility
Language:	English, Norwegian, Swedish, Danish, Icelandic, Persian and Spanish
Study design:	Original studies (sensitivity, specificity etc.), systematic reviews, and non-systematic reviews (feasibility)

Exclusion criteria

Study design:	Case-reports, case series, animal studies and studies not available in full text (e.g., conference abstracts)
Population:	Patients with non-solid tumours (e.g., leukaemia)
Intervention/ comparator:	Other tests than those listed above
Outcomes:	Outcomes not related to the test accuracy or to the feasibility of tests

Other exclusion criteria were studies that did not report a comparison between tests, or with cell-lines with known mutation status, or studies written in other languages than those listed above.

Selection of studies

We downloaded all titles and abstracts retrieved by the electronic searches into the reference management program EndNote (37) and removed duplicates. The references were then exported to Rayyan (38) for screening. Two review authors (GMF and JFME) independently assessed titles and abstracts against the inclusion criteria. We

obtained full-text copies of potentially relevant studies and assessed them in duplicate. We resolved disagreements by discussion. Reasons for exclusion of publications read in full text are reported in Appendix 4.

Data extraction and management

One reviewer (GMF or JFME) independently extracted data from each included study into a standardised and piloted data extraction form, which was adapted for use in this HTA. The other reviewer quality-checked the accuracy of the data extraction. Any disagreements were resolved through discussion among review authors. We extracted the following data: citation, year of publication, setting, country, funding, conflicts of interest, study designs, language, and details on the PICOs:

- *Participants*: number, age, gender, race/ethnicity, socioeconomic status, time since diagnosis, previous treatment received, concomitant therapy/medication, etc.
- *Molecular tests*: technical details of tests, regulatory status, in-house or commercial test, previous tests conducted, sequence of tests if more than one test, test turnaround time (TAT), type and amount of biological tissue required for the tests, etc.
- *Comparisons*: head-to-head-comparisons, index test(s) versus reference tests if applicable, or cell-lines with known mutation status
- *Outcomes*: analytical validity (sensitivity, specificity, assay robustness, quality control), clinical validity (i.e., sensitivity and specificity, positive and negative predictive values), clinical utility (e.g., response rate to treatment, overall survival, quality of life), advantages and limitations of the different tests etc.

We also, when needed, contacted authors over e-mail for clarification of results. One reminder was sent after approximately one week if no response was received to the first e-mail. If no response was received, and we were unable to resolve the issue, the study was excluded from the review.

Quality assessment

Two authors (GMF and JFME) used the three-step process suggested by the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) working group (32) to assess the quality of the evidence of included original studies using what the EGAPP-group refers to as a 'chain of evidence'. See Appendix 5. The three steps include:

(i) Determining the hierarchy of the data source and study design (Level 1 to 4, of which level one is the highest) for the components of the evaluation (i.e., analytical validity, clinical validity, and clinical utility);

(ii) Determining the quality of individual studies (internal validity), and,
(iii) Grading the quality of evidence for the individual components of the chain of evidence (convincing, adequate, or insufficient).

We did not quality assess the included narrative review, or the expert opinion papers, as these, by nature are considered being of low quality (37).

In addition, we used the Standards for Reporting of Diagnostic Accuracy (STARD) checklist (39), which contains 30 essential items, to assess the quality of the reporting in the included original studies. One author (GMF) assessed the quality of reporting, and a second author (JFME) double-checked the assessment. Any discrepancies were solved though discussion between the two authors.

Compilation of results

Meta-analysis was not feasible as studies were heterogeneous in terms of study populations (type of cancer, and thus biomarker prevalence), comparator (reference) test, cut-off criteria for positivity, type of accuracy related outcome reported etc. We have provided a narrative summary of the available evidence from original studies on the test accuracy of different analytical techniques for the detection of NTRK gene fusions in text and tables. In addition, data on the characteristics (e.g., advantages, and limitations) of these tests retrieved from the narrative reviews and expert opinion papers have been summarised in text and tables. Additional data is provided in appendices.

Review results

Search results

See Figure 1. PRISMA study flow chart (40).

The search of the electronic databases yielded 2,379 unique citations after removing 707 duplicates. Two-thousand three-hundred and fifty-six of these were irrelevant and directly excluded at the title and abstract screening stage, leaving 22 citations to be retrieved in full text for further scrutiny. Two studies with unclear results were excluded when clarifications requested from the authors were not received (41;42). Nine original studies (11;43-50) that provided data on comparisons between tests for the detection of NTRK gene fusions in solid tumours, were included in this HTA. In addition, we included five narrative reviews (19;51-54), and two expert opinion papers (18;55) to address the advantages and limitations of the tests. Studies read in full text but subsequently excluded (N=8) are listed in Appendix 4, along with the reasons for exclusion.

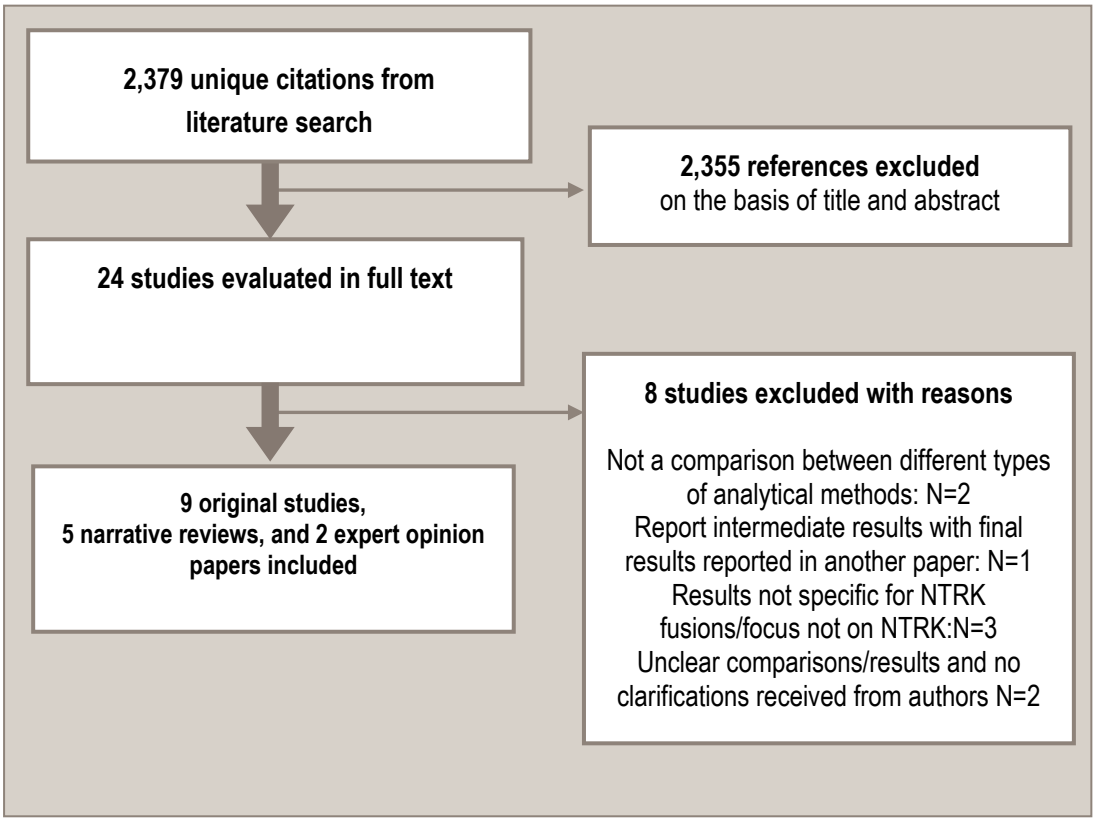

***Figure 1**. PRISMA study flow chart (40)*

Characteristics of included original studies

See Table 3 Prevalence of NTRK gene fusions, Table 4 Test comparisons, Table 5 Characteristics of included studies, and Appendix 6 Technical details of included tests.

Study design, and country of origin

We included nine original studies (11;43-50). Eight studies conducted retrospective analyses of stored data, and in one study it was unclear whether it was prospective or retrospective (43). Four of the nine studies were conducted in the USA (11;43;48;50); and one in Austria (44), Romania (47), Sweden (46), Korea (45), and Taiwan (49) respectively.

Population: Selection of participants and their characteristics

Bell and colleagues (43) assessed salivary gland carcinomas samples (N=70) that included a number of different subtypes (19 secretory carcinomas, 43 acinic cell carcinomas, 3 salivary hybrid carcinomas, 2 mucoepidermoid carcinomas, 1 salivary duct carcinoma, and 2 sinonasal adenocarcinomas), and different tumour sites (parotid gland (n=58), parapharyngeal space (n=2), neck lymph node (n=2), and other locations (n=8). No information was provided on the recruitment/ selection of patients into the study, nor were any patient (sample) characteristics provided.

Bricic et al. (44) included 494 soft tissue sarcomas (26 different types) that either had been diagnosed between 1999 and 2019 at the Diagnostic and Research Institute of Pathology, Medical University of Graz, or selected cases that had been diagnosed between 2017 and 2019 that were identified from the consultation files of one of the authors. The age of NTRK positive patients ranged from eight months to 50 years.

Choi et al. (45) included a cohort of 80 patients with T3 or T4 colorectal carcinomas (44 men and 36 women), with a mean age of 67.8 years (range, 31-87 years) that underwent resection of the primary tumour at Pusan National University Hospital (PNUH) between January and May 2015. FFPE sections were obtained from the Department of Pathology and the National Biobank of Korea, Pusan National University Hospital.

Elfving and colleagues (46) included 688 primary NSCLC patients (617 samples were evaluable), from two patient cohorts representative for the operable Swedish NSCLC population: one cohort (Uppsala I) which included 360 patients operated 1995–2005, and the other (Uppsala II) which included 328 patients operated 2006–2010.

Fu et al. (47) included a cohort of 819 unselected patients who underwent surgical resection for CRC identified by searching the database of the Department of Pathology, Nanjing Drum Tower Hospital, Nanjing, China for all cases between 2015 and 2020. The inclusion criteria were as follows: 1) pathologically diagnosed adenocarcinoma, mucinous adenocarcinoma, or high-grade neoplasia according to the latest WHO classification; 2) complete clinical and pathological data. Exclusion criteria included: 1) extracolonic and appendiceal location; 2) tumours undergoing biopsy alone or treated endo-luminally; 3) preoperative local or systematic anticancer neoadjuvant

therapy; or 4) incomplete clinical data. Demographic and clinicopathological data were reported for both NTRK positive and negative cases.

Gatalica et al (48) included various solid cancer types profiled at a commercial laboratory in the United States (Caris Life Sciences) from 2015 to 2018. Studied cancers with NTRK fusions included NSCLC (n =4073), colorectal carcinomas (n = 1272), gliomas (n =982), breast carcinomas (n = 769), various soft tissue sarcomas (n = 478), cancers of unknown primary (n = 227), thyroid carcinomas (n = 70), cervical carcinomas (n = 68). The patients were mainly adults with a mean age of 55 years, with the exception of one 11-year-old male patient with glioblastoma multiforme. The authors included cancers of various other primary cancer sites without NTRK fusions for comparison (>3000).

Lee et al (49) identified 525 consecutive papillary thyroid carcinoma cases from the pathology archives of Taipei Veterans General Hospital between October 2015 and March 2019. The patients had a mean age of 49.7 years (range: 11–86 years) and all had a tumour size larger than 0.5 cm at diagnosis. Sixty of the 525 patients who previously had tested IHC BRAF negative were included in the study.

Rudzinski and colleagues (50) included 60 patients with diagnoses of infantile fibrosarcoma, congenital cellular mesoblastic nephroma, or patients for which these two were considered in the differential diagnosis, from the pathology databases at Seattle Children's Hospital and University of California San Francisco. Samples (N=49) that had not previously been tested with FISH or RT-PCR (i.e., with confirmed ETV6/ETV6-NTRK3 fusions; N=7), were submitted for NGS (n=49). Four additional cases with confirmed NTRK rearrangements were included from other institutions. Paediatric mesenchymal tumours not harbouring NTRK fusions (n=28), and miscellaneous soft tissue tumours, which were well-defined histologically, and not expected to harbour NTRK fusions (n=22) were included as negative controls. No information on type and age of the samples were provided.

Salomon et al. (11) was a retrospective review of the MSK-IMPACT (DNA sequencing) and MSK-Fusion panel (RNA sequencing) results from January 1st, 2014 to March 30th, 2019. All 38,095 tumour samples from 33,997 patients were FFPE, and all testing had been performed in CLIA approved laboratories and reported clinically. Tested by pan-Trk IHC were a total of 66 fusion positive cases and 317 fusion negative cases. Thirteen various cancer types were included (Salivary gland carcinoma (13); thyroid carcinoma (13); sarcoma (13); lung adenocarcinoma (9); colorectal carcinoma (9); glioma/neuroepithelial tumour (8); breast carcinoma (6), pancreatic adenocarcinoma (5): melanoma (4); inflammatory myofibroblastic tumour (3); cholangiocarcinoma (2); appendiceal adenocarcinoma (1), and neuroendocrine tumour (1)). No other patient characteristics was reported. The reported prevalence of NTRK fusions in twelve of these tumours was low: below 1 % in 10 cancers, and between 2-5% in two tumour types (i.e., thyroid carcinoma and salivary gland carcinoma). In inflammatory myofibroblastic tumours the prevalence of NTRK fusions was 17.7%.

Prevalence of NTRK fusions across cancer types

The prevalence of NTRK fusions varied from <1% and up to 90% across the various types of solid tumours assessed in the included studies (16;17). Due to the suggested importance of biomarker prevalence for the test accuracy (18;19), we categorised the tumours assessed in the included studies as low frequency (<5% NTRK gene fusions), intermediate frequency (5-<25%), and high frequency (>80 %) according to what has been done by others (18;19). Four studies (44-47) included cancers with low (<5%) frequency of NTRK fusions (CRC, NSCLC, and STS). Salomon et al. included 11 types of low frequency carcinomas (see footnotes table 3 for details), and two cancer types with intermediate (5-25%) frequency (salivary gland carcinomas, inflammatory myofibroblastic tumours). It should be noted that the prevalence of NTRK fusions in many of these tumours was much higher than otherwise reported in the literature. Gatalica and colleagues (48) included seven low frequency tumours (see footnotes table 3), and one intermediate frequency tumour (Thyroid carcinoma). Lee et al (49) included one intermediate frequency carcinoma (PTC). Bell et al and Rudzinski et al (43;50) included one high frequency tumour each (infantile fibrosarcoma and secretory carcinomas of the salivary gland respectively), and one intermediate frequency cancer (congenital mesoblastic nephroma and acinic carcinomas respectively), and one cancer with unknown frequency (hybrid carcinomas).

Four studies provided accuracy-related test results for colorectal carcinomas (11;45;47;48); three studies reported results for NSCLC/lung cancer (11;46;48), for PTC/Thyroid carcinomas (11;48;49), for salivary gland carcinomas (11;43;48), and two studies reported results for soft tissue sarcomas (44;48). For 11 types of solid tumours only single studies provided accuracy-related data.

Table 3. *Prevalence of NTRK fusions across different tumour types*

	Colorectal carcinomas (CRC) **(45;47)**	**Non-small cell lung cancer** (NSCLC) **(46)**	**Papillary Thyroid Carcinoma** (PTC) **(49)**	**Paediatric mesenchymal tumours** (infantile fibrosarcoma (IFS), and congenital mesoblastic nephroma, (CMN)) **(50)**	**Salivary gland cancers** (secretory carcinomas, acinic carcinomas, and hybrid carcinomas) **(43)**	**Soft tissue sarcomas** (STS) **(44)**	**Various cancer types** [a, b] **(11;48)**
NTRK fusion	NTRK1, 3	NTRK1, 2	NTRK 1,3	NTRK	ETV6-NTRK3	NTRK1	NTRK 1-3
Prevalence reported in the literature	0.22%-0.26% (16;17)	0.17%-0.24% (16;17)	5-25% (adults) (19); 25.93% (paediatric patients (17)	>80% (ESMO 2019); IFS: 90,6% and CMN:21.5% (17)	>80% in secretory carcinomas (19) (79.68% in (17)); 11.1% in acinic cell carcinomas (17); unclear prevalence in hybrid carcinomas	1.27% (adults); 4.02% (paediatric patients) (16)	1.60%(16) to 5-25%(19) thyroid carcinomas;2.43% salivary gland carcinomas (16), and 17.7% Myofibroblastic tumours (11)
Frequency	Low	Low	Intermediate	High (1) Intermediate (1)	High (1), Intermediate (1); unclear (1)	Low	Low (11+7), Intermediate (3),

a Salomon 2020: 10 low-frequency carcinomas (most of them <1%): lung, pancreas, biliary tract, appendix, colorectal, sarcomas, gliomas/neuroepithelial tumours, breast carcinomas, melanomas, and neuroendocrine tumours; Intermediate frequency carcinomas: thyroid carcinomas, and inflammatory myofibroblastic tumours; it should be noted that the prevalence of NTRK fusions in many of the included cancer types had much higher prevalence than what have been reported in the literature. b Gatalica 2019: 7 low frequency carcinomas: NSCLC, colorectal carcinomas, gliomas, breast carcinomas, various soft tissue sarcomas, cancers of unknown primary origin, cervical carcinomas; Intermediate frequency: thyroid carcinomas; salivary gland carcinomas

Intervention (index) test/Comparator (reference) tests

IHC was in all nine studies compared with one or more analytical technique for detection of one or more NTRK fusion. The comparator tests were as follows: FISH (43;45;47;49); RT-PCR (and to some extent FISH) (43); RNA-based NGS (44;46;48;49); DNA-based NGS (47;50); DNA and RNA-based NGS (11). Two studies compared results of IHC and/or FISH with NGS (47;49). One compared DNA-based NGS with RNA-based NGS (11).

The number of samples tested by at least two different analytical techniques ranged from 15 (43) to 4,136 (48) across studies (median: 60 samples). In most studies, it was unclear if the number of samples equalled the number of participants, or if some participants contributed more than one sample. In three studies, all samples were tested with more than one method (45;47;49), while in four studies, only samples that tested positive for NTRK fusions with one method (or a subset of samples with sufficient

material), were tested with one or more other methods (See Table 4). In three studies a selection of positive and negative samples (typically verified with RNA- and/or DNA-based NGS) were tested with the index test (IHC) (11;48;50).

Table 4. *Test comparisons and no of samples tested in the included studies (N=9)*

Author Year	IHC	FISH	RT-PCR	DNA- NGS	RNA-NGS
Bell 2020 (43)	70	15 (as part of routine care)	45 (with sufficient material)	-	-
Bricic 2021 (44)	494	-	-		16 (IHC positive cases)
Choi 2018 (45)	80	80	-	-	-
Elfving 2021 (46)	617				11 (IHC positive cases)
Fu 2021 (47)	819	819		18 (IHC or FISH pos. cases)	18 (IHC or FISH positive cases)
Gatalica 2019 (48)	4,136 (28 positive cases)				11, 502
Lee 2020 (49)	60	60			6 (sub-sample of 12 FISH positive cases)
Rudzinski 2018 (50)	79 (28 negative cases and 22 cases not expected to harbour NTRK fusions)	-	-	49 (subsample)	-
Salomon 2019a (11)	66 positive and 317 negative cases	-	-	38,095	2,189

DNA: Dioxiribonucleic acid; FISH: Fluorescence in situ Hybridisation; IHC: Immunohistochemistry; NGS: Next Generation Sequencing; RNA: Ribonucleic acid; RT-PCR: reverse transcriptome polymerase chain reaction

Outcomes

The NTRK gene fusions detected in the included studies were as follows: ETV6-NTRK3 (43); NTRK 1 (45); NTRK1 and NTRK 3 with any fusion partner (44). Six studies included any NTRK fusion, and two of these studies reported results for the three NTRK fusions separately (11;46;48).

Six studies reported on sensitivity and specificity of (index) tests (11;43;47-50). Two studies reported concordance between test results (43;48). Three studies reported consistency between (positive) test results (44;46;47). One study reported area under

the receiver operating characteristics curve (AUROC) (45). Only two studies provided a point estimate, with a measure of dispersion (43;45).

Test and sample characteristics

Sample types

All studies used FFPE tissue samples, and in three studies FFPE was used to construct tissue micro arrays (TMAs) for analysis (44;46;49). The age of samples (duration of storage) varied across included studies from a couple of years and up to 26 years (46). In one study the age of the samples was unclear (43).

IHC Antibody clones, FISH probes, and NGS systems

Seven studies used the Pan-Trk rabbit monoclonal antibody, clone EPR17341 from Abcam (11;43;46;48-50), but with different dilutions (between 1:125-1:50), or the same type of clone but from Roche (44). Two studies used other type of clones.

Three studies (45;47;49) used dual-colour break-apart FISH probes from different suppliers. Six studies used NGS systems from different suppliers, that were either RNA-based (44;46;49), both DNA- and RNA-based (11;47), or mostly DNA-based (50). The systems could be either amplicon- or hybridization-capture based (See Appendix 6 for details).

Cut-off criteria/standard for test positivity

The cut-off criteria for positivity for IHC varied across included studies, and two studies (45;50) provided no clear criteria. Also, for FISH did the cut-off criteria vary somewhat across studies. One study (43) stated that the normal cut-offs were established in the lab, but the actual criteria were not described. No cut-off criteria /standards were reported for RT-PCR or NGS in any of the included studies.

Table 5. *Characteristics of included original studies (N=9)*

	Bell 2020 (43)	Brčić 2020 (44)	Choi 2018 (45)	Elfving 2021 (46)	Fu 2021 (47)	Gatalica 2019 (48)	Lee 2020 (49)	Rudzinski 2018 (50)	Solomon 2019a (11)
Country	USA	Austria	Korea	Sweden	Romania	USA	Taiwan	USA	USA
Aims	To determine whether pan-Trk-IHC could detect ETV6-NTRK3 fusions as *reliably* as RT-PCR and FISH.	To gain further insights into the staining profile with the pan-TRK assay, and correlate our findings with molecular testing	To investigate the potential use of IHC for detecting NTRK1 gene fusions, a comparison with FISH	To evaluate the newly introduced diagnostic immunohistochemical assay (clone EPR17341) on a representative NSCLC cohort	To gain insight into the clinicopathologic profile of CRC harbouring oncogenic NTRK fusions based on eastern populations as well as make the best testing algorithm for the screen	To review a large cohort of solid malignancies profiled by a commercial laboratory tested for NTRK gene fusions and other pathogenic/targetable genomic and protein alterations. To assess diagnostic utility of immunohistochemistry in detecting NTRK gene fusions	To characterize the clinic-pathological features of PTC with NTRK1/3 fusions, to examine the utility of pan-TRK IHC, and to compare IHC with FISH and NGS.	To evaluate the performance of IHC staining using pan-Trk and TrkA antibodies	To investigate the performance of IHC and DNA-based NGS to indirectly or directly detect NTRK fusions relative to an RNA-based NGS approach
Cancer type	Secretory carcinoma (SC) of the salivary gland	Paediatric soft tissue sarcomas	Colorectal carcinoma (CRC):	Non-Small Cell Lung Cancer (NSCLC)	Colorectal carcinomas (CRC)	Various solid cancer type (mostly carcinomas and brain gliomas)	Papillary Thyroid Carcinoma (PTC)	Paediatric mesenchymal tumours (PMT)	13 cancer types
Gene fusions	ETV6-NTRK3	NTRK 1/3 with various fusion partners	NTRK1	NTRK fusions (any)	NTRK fusions (any)	NTRK fusions (any)	NTRK 1/3	NTRK fusions (any)	NTRK 1/2/3,

No of pts.	70	494	80	688	819	11,502	60	79	33,997 (38,095 samples)
Characteristics of pts.	NR	NR	44 men and 36 women; mean age of 67.8 yrs. (range, 31-87 yrs.); CRT T3 or T4	2 cohorts; 1995-2005, and 2006-2010; Patients aged 40 to 84 yrs (median 67yrs); 52.5% females and 47.5% males	NR; large, unselected cohort	Adult patients and one paediatric patient with gliobastoma multiforme, period 2015-2018	$BRAF_{V600E}$-negative cases from a cohort of 525 consecutive cases, none of which had previously received radiation	28 negative controls, and 22 people not expected to harbour NTRK fusions	NR
IHC antibody clone	Pan-Trk antibody (Clone: EPR17341, Abcam, USA), (dilution, 1:125;),	Antibody (clone EPR17341, RTU, Roche/Ventana)	Anti-TrkA C-terminal monoclonal antibody	Pan-Trk antibody (Clone: EPR17341, Abcam, USA),	Pan-Trk antibody (Clone: EPR17341, Abcam, USA); and others	Pan-Trk antibody (Clone: EPR17341, Abcam, USA);	Pan-Trk antibody (Clone: EPR17341, Abcam, USA),	Pan-Trk antibody (Clone: EPR17341, Abcam, USA),	Pan-Trk antibody (Clone: EPR17341, Abcam, USA),
Criteria for positivity	IHC: Nuclear, cytoplasmic, or membranous staining in more than 5% of tumour cells RT-PCR: -	IHC: Any cytoplasmic/ unclear staining in more than 1% of tumour cells NGS: -	IHC: No cut-off criteria for positivity provided. Semi-quantitative scoring system used. FISH: when more than 20 out of 100 nuclei demonstrated break-apart 5'- and 3'-end signals	IHC: Staining intensity of moderate or strong in ≥1 % of tumour cells NGS: -	IHC: Cytoplasmic staining intensity was considered positive FISH: 15% break-apart signals, or the same percentage with single green/red signals NGS:-	IHC: positive if ≥1% of tumour cells exhibited positivity at any intensity above background.	IHC: Any unequivocal immunoreactivity on cytoplasm and/or nuclei with clear contrast with surrounding non-tumorous tissue FISH: more than 20% of nuclei positive for break-apart signals, namely two separate green and orange signals with distance more than one signal diameter	IHC: No cut-off criteria for positivity provided. NGS: -	IHC: Staining above background in at least 1% of tumour cells in any pattern including membranous, cytoplasmic, perinuclear, or nuclear RNA-/DNA-NGS: NR

Intervention (index) test	Pan Trk.IHC	Pan-Trk IHC	Pan Trk IHC	Pan Trk IHC	IHC	Pan Trk IHC	Pan Trk IHC	PanTrk-IHC	Pan-Trk-IHC, and DNA-NGS
Comparator (reference)	RT-PCR and FISH	RNA-NGS	FISH	RNA-NGS	FISH and NGS	RNA-NGS	FISH and NGS	DNA-NGS (various systems/ platforms)	RNA- NGS
Outcomes r	Sensitivity, specificity, and concordance	Consistency	AUROC, consistency	Consistency	Sensitivity, specificity (calculated by review authors)	Sensitivity, specificity, and concordance	Sensitivity, specificity (calculated by review authors)	Sensitivity, specific-ity	Sensitivity, specific-ity, PPV, NPV

AUROC: Area Under the Operating Received Characteristics Curve; BRAF: proto-oncogene B-Raf*; DNA-NGS: Deoxyribonucleic acid NGS; FISH: Fluorescence In Situ Hybridisation; IHC: ImmunoHistoChemistry; NGS: Next Generation Sequencing; NPV: Negative Predictive Value; NR: Not Reported; NTRK: add; PPV: Positive Predictive value; RNA-NGS: Ribonucleic acid NGS; RT-PCR: Reverse Transcriptome Polymerase Chain Reaction; TMA: Tissue Micro Array*

Results- outcomes related to test accuracy

See Table 6 Results related to test accuracy

Four studies compared IHC with FISH for the detection of NTRK gene fusions; one study compared IHC (43;45;47;50) with RT-PCR (43); six studies compared IHC with RNA-NGS (11;44;46-48), two compared IHC with DNA-NGS (47;50), and one study also compared DNA- with RNA-based NGS (11). In addition, one study compared IHC/FISH positive results with DNA- and RNA-NGS (47).

IHC vs. FISH (4 studies)

Sensitivity and specificity: The sensitivity of Pan Trk IHC ranged from 38.5% to 41.7 % across two studies, and the specificity from 99.4% to 100% for detecting *any NTRK gene fusions* in colorectal carcinomas (47) and in BRAFV600E-negative papillary thyroid carcinomas respectively (49). Positive predictive value (PPV) ranged from 33% (47) to 100% (49) and negative predictive value (NPV) from 99% (47) to 87.3% (49).

AUROC: One study (45) reported a significant correlation between the IHC and FISH results for detection of *NTRK1 gene fusions* in colorectal carcinomas, and an area under the receiver operating characteristic (ROC) curve of 0.926 (0.864-0.987, 95% CI, P = .001). It should be noted that while FISH was reported to be positive in 6/80 samples (7.8%), the number of IHC positive cases was unclear as no cut-off criteria for positivity were provided.

Concordance: One study (43) reported fair agreement (concordance) between IHC and FISH for the detection of *ETV6-NTRK3 gene fusions* in secretory carcinomas of the salivary gland (0.359; SE:0.218 (95% CI:0, 0.786), However, only a small subsample were included in this analysis.

IHC vs. RT-PCR (1 study)

Sensitivity and specificity: One study (43) reported 90,9% (10/11) sensitivity and 100% (34/34) specificity of PanTrk IHC in detecting *ETV6-NTRK3 gene fusions* in secretory carcinomas of the salivary gland.

Concordance: The same study (43) also reported almost perfect concordance between test results (0.938 (SE:0.061; 95%CI: 0.818; 1).

IHC vs. RNA-NGS (6 studies)

Sensitivity and specificity: Salomon et al (11) reported 96.2% sensitivity of Pan-Trk IHC for detecting *NTRK1 fusions*, 100% for *NTRK2 fusions*, and 79.4% for detecting *NTRK3 fusions* (total sensitivity:87.9%). Total specificity across all 13 cancer types was 81.1%. Sensitivity varied across type of cancer from 80% in breast cancer to 100% in other cancer types (e.g., inflammatory myofibroblastic tumour, appendix, glioma, cholangio, melanoma). Specificity also varied from 20.8% in gliomas to 100% for other cancers (e.g., inflammatory myofibroblastic tumour, colon, lung, thyroid cancer, appendix, cholangio, melanoma).

Gatalica (48) reported overall sensitivity of 75% and specificity of 95.9% of IHC in detecting NTRK gene fusions across a number of different cancer types, and a PPV and NPV of 11.2% and 99.8%, respectively. Gatalica also reported on the concordance between the results of RNA-based NGS and IHC; 87.5% (7 of 8 cases) NTRK1 fusion positive with NGS were also positive with IHC, 88.7% (8 of 9) NTRK2 fusion positive cases with NGS were also positive with IHC, and so was 54.5% (6 of 11) NTRK3 positive cases. IHC was positive in 4.5% of samples (187 of 4126) and RNA-NGS was positive in 0.27% of samples (31/4136), suggesting that 16.6% of samples were concordant. This study did not report sensitivity and specificity for the different cancer types separately.

Consistency: Three studies (44;46;47) reported that of samples positive with IHC the proportion of consistent positive samples (46) with RNA-NGS were 0% for NSCLC samples (44), 31% for paediatric soft tissue sarcoma samples, and 60% for CRC samples (47).

IHC vs. DNA-NGS (2 studies)

Sensitivity and specificity: One study (50) reported a 97% sensitivity and 98% specificity of Pan-Trk IHC, and 100% sensitivity and 63% specificity for TrkA IHC in detecting any NTRK rearrangements in paediatric mesenchymal tumours. Various (mostly) DNA-based NGS systems constituted the reference standard.

Consistency: One study (47) reported that 30% of samples found NTRK positive with IHC, were also positive (consistent) with DNA-NGS.

IHC/FISH vs. DNA-NGS or RNA- NGS (1 study)

Consistency: One study (47) reported that 17% of IHC and/or FISH positive samples, also were positive (consistent) with DNA-based NGS, and that 72% of samples positive with IHC and/or FISH also were positive (consistent) with RNA-NGS.

DNA-NGS vs. RNA-NGS (1 study)

Sensitivity and specificity: One study (11) reported an overall sensitivity of DNA based NGS (MSK-IMPACT) of 96.8% for detecting NTRK1 fusions, 0% for NTRK2 fusions, and 76.9% for detecting NTRK3 fusions as compared to RNA-based NGS. Total sensitivity was 81.1%, and total specificity across the 13 cancer types was 99.86%.

Accuracy related outcomes by cancer types were presented in Appendix 7.

Table 6. Summary of accuracy related results from included studies(N=9)

Author Year	Fusion	Sensitivity	Specificity	PPV	NPV	AUROC	Concordance (Cohen's kappa)	Consistency (positive samples)
IHC vs. FISH								
Bell 2020[1] (43)	ETV6-NTRK3	-	-	-	-	-	0.359 (SE:0.218; 95%CI: 0; 0.786)	-
Choi 2018[2] (45)	NTRK1	-	-	-	-	0.926 (range: 0.864-0.987)	-	-
Fu 2021[3] (47)	NTRK (any)	38.5% (5/13)	99.4%	-	-	-	-	-
Lee 2020[4] (49)	NTRK (any)	41.7%	100%	100%	87.3%	-	-	-
IHC vs. RT-PCR								
Bell 2020 (43)	ETV6-NTRK3	90.9% (10/11)	100% (34/34)	-	-	-	0.938 (SE:0.061; 95%CI: 0.818; 1)	-
IHC vs. RNA-based NGS								
Bricic 2020 (44)	NTRK1, NTRK3	-	-	-	-	-	-	IHC:16/494; RNA-NGS: 5/16 (31.2%)
Elfving 2021(46)	NTRK (any)	-	-	-	-	-	-	IHC:11/ 617; RNA-NGS: 0 /11 (0%)
Fu 2021 (47)	NTRK (any)	-	-	-	-	-	-	IHC: 10/819; RNA-NGS: 6/10 (60%)
Gatalica 2019 (48)	NTRK (any)	75%	95.9%				IHC 187/4126 (4.5%); RNA NGS 31/4136 (0.27%)	
Solomon 2019[5] (11)	NTRK1, NTRK2, and NTRK3	NTRK1: 96.2% (26/27) NTRK2: 100% (5/5) NTRK3: 79.4% (27/34) Overall: 87.9% (58/66)	81.1%. (257/317)	-	-	-	-	-
IHC vs. DNA-based NGS								

Fu 2021 (47)	NTRK (any)	-	-	-	-	-	-	IHC: 10 /494; DNA-NGS: 3 /10 (30%)
Rudzinski 2018[6] (50)	NTRK (any)	Pan Trk: 96.7%, TrkA: 100%	Pan Trk: 97.9%; TrkA:63.3%	PanTrk: 96.7%; TrkA:59.1%	Pan Trk: 97.9%; TrkA:100%.	-	-	-
FISH vs. RNA-based NGS								
Fu 2021 (47)	NTRK (any)	-	-	-	-	-	-	FISH 13/819: RNA-NGS: 12/13 (92.3%)
FISH vs. DNA-based NGS								
Fu 2021 (47)	NTRK (any)	-	-	-	-	-	-	FISH: 13/819: DNA-NGS: 3/13 (23.1%)
DNA-based NGS vs. RNA-based NGS								
Salomon 2019a (11)	NTRK1, NTRK2, and NTRK3	NTRK1 96.8% (30/31); NTRK2 0% (0/4); NTRK3 76.9% (30/39); Total 81.1% (60/74)	99.86% (33877/ 33923)	-	-	-	-	-

DNA: Deoxyribonucleic acid; IHC: Immunohistochemistry; FISH: Fluorescence In situ Hybridization; NPV: negative predictive value; NGS: Next Generation Sequencing; RNA: RT-PCR: Reverse Transcriptome Polymerase Chain Reaction; PPV: positive predictive value; ROC-curve: Receiver operating characteristic curve ; 1 Bell 2020: 15 of 70 samples were tested with both IHC and FISH (as part of routine care); 45 of 70 samples (with sufficient material) were tested with IHC and RT-PCRs; 2 Choi 2018: used TrkA immunostaining; 3 Fu 2021: Results calculated by review authors; 4 Lee 2020: only a subset of positive FISH cases (6 of 12) were tested with RNA-based NGS; 5 Salomon 2019a: details on the different types of cancer can be found in the original paper; 6 Rudzinski 2018: Most analyses by DNA based NGS (a local system that according to personal communication with the authors may not have picked up on all fusions), but not clear how many were analysed with RNA-based NGS (communication with authors); TrkA immunostaining was used in addition to PanTrk IHC.

Test algorithms suggested in included studies

Four of the included studies proposed a testing algorithm for the detection of NTRK fusions (11;47;49;50). Apart from Solomon 2020 (11) all testing algorithms propose Pan-TRK IHC as the first screening tool for detecting NTRK fusions that can be further confirmed by either FISH or NGS. The testing algorithms had some particularities:

- Fu and colleagues (47) recommended the use of both FISH and DNA mismatch repair (MMR) to confirm classical NTRK fusions due to the poor ability of FISH to identify classical, sub classical, and non-classical fusions in CRC (low NTRK frequency). NGS is recommended as a confirmatory test when sub classical NTRK fusions are identified with Pan-TRK IHC and FISH.
- Lee et al (49) proposed to first triage PTC cases (intermediate NTRK frequency) based on BRAFV600E status, followed by pan TRK IHC in BRAFV600E-negative cases (as positive cases are very unlikely to harour NTRK fusions). IHC NTRK fusion positive histology may be a useful tool for cases negative to both biomarkers. Further molecular testing (e.g., with FISH, NGS) may be consisdered for cases showing any suggestive morphologic features, non-infiltrative tumour border, clear cell change, and reduced nuclear elongation and irregularity). The presence of NTRK fusions is unlikely in samples without these morphological features.
- Rudzinski et al (50) proposed Pan-TRK IHC as a diagnostic surrogate for NTRK rearrangements in presence of moderate to strong cytoplasmatic staining in pediatric mesenchymal tumours (high NTRK frequency), and NGS for weak cytoplasmatic staining.
- Solomon and colleagues (11) recommended the use of RNA-based NGS as the first diagnostic test to detect NTRK fusions in sarcomas (low NTRK frequency). Authors acknowledged Pan-TRK IHC performed well in carcinomas of colon, thyroid, pancreas, and lung.

All the studies highlighted the advantages of Pan-TRK IHC as a diagnostic tool i.e., that IHC is cheap, feasible, widely available, and shows high sensitivity and specificity. The studies identified NGS as costly and of limited availability.

Quality of included evidence: results of the EGAPP tool

See Appendix 8. Quality of included evidence: results of the EGAPP tool assessment. The EGAPP tool items are described in Appendix 5.

The results of the 3-steps EGAPP quality assessment are described below. The tool, as mentioned earlier, involves assessment of analytical validity, clinical validity, and clinical utility. None of the included studies had as the main aim to assess the analyt-

ical validity in the laboratory (technical performance), nor did any of the studies report data on clinical utility (risks and benefits for the patients). We have summarised results related to clinical validity below.

***Step 1**. Assessment of the study hierarchy (level 1-4, with 1 being the highest):*
We judged eight of the nine included studies to be level 3 in the hierarchy of study level evidence EGAPP proposes (lower quality case-control or cross-sectional studies). One study (11) was judged to be a level 2 study (well-designed case-control study).

***Step 2**. Assessment of the internal validity of included studies:*
Clear description of the disorder/phenotype and outcomes of interest. Status (biomarker positive or negative) was confirmed for all cases in three studies (43;45;49). In six studies (11;44;46-48;50) a subgroup of cases (positive with index test, or samples with sufficient material) were assessed with the reference test. Three of these studies also included negative control samples (11;48;50). In one study it was unclear how some of the negative samples were confirmed, and some were expected not to harbour NTRK fusions due to their histology (50). In one study (11) positive samples were positive with either DNA- or RNA based NGS, and negative samples were either positive with DNA-NGS but negative with RNA-NGS, or just negative with RNA-NGS. In one study authors mentioned the inclusion of 'confirmed' positive samples (48). Only one of the studies (11) provided an appropriate description of included control material. Three of the nine studies reported on the biomarker prevalence in the population of interest (46;47;49).

Adequate description of study design and test/methodology. As in the STARD assessment, we found that the included studies lacked a description of the study design, and that even though all studies described the tests under study, the level of detail may not have been sufficient to allow replication. Two studies described the statistical methods used to calculate the accuracy outcomes (43).

Adequate description of the study population. Only one of the nine studies provided inclusion and exclusion criteria (47). The study populations were poorly defined in most studies, and only in three studies were the study samples suggested to be representative for the clinical population (46-48). In some of the other studies (11;49) the study population might not have been representative. None of the included studies reported whether allele/genotype frequencies or analyte distributions were known in general and subpopulations.

Independent blind comparison with appropriate, credible reference standard(s). Three studies applied the reference test to all samples (45;47;49). Few studies reported how indeterminate results were managed for the different tests, and none described how missing data/data of poor quality for verification with NGS was handled (46). None of the included studies provided any information on whether those conducting the index test(s) were blinded to the results of the reference test(s) and vice versa.

Analysis of data. Possible biases were identified, and their potential impact were at least partly discussed in five of the studies (11;45-47;49). Point estimates of sensitivity

and specificity (and ROC-curve) were only provided in six studies, but only in two studies was a measure of dispersion provided. Estimates of positive and negative predictive value were reported in four studies and for some of the cancers under investigation.

Step 3. Grading of the certainty of evidence (convincing- adequate- inadequate)

We judged that eight of the included studies provided inadequate evidence for clinical validity. One study (11) was judged to provide (close to) adequate evidence. For no study was the evidence 'convincing'.

Quality of reporting–results of the STARD checklist

The results of the STARD checklist revealed overall poor reporting across included studies. The results of the 30 checklist items are summarised in the text below and presented in detail in Appendix 9.

Item 1-4. Title, abstract, background. None of the nine included studies self-identified as a test accuracy study in the title, but it was indicated in the abstract. Only one study provided a structured abstract describing methods, results, and conclusions (46), but no study described the study design. All studies provided a scientific and clinical background, with a description of relevant tests, and at least some explanation (objectives) as to why the study had been conducted. None of the studies provided a study hypothesis.

Item 5. Study design. Eight studies were retrospective, i.e., data collection was not planned before the index test and/or reference standard were performed (eligible samples were identified from registers). In one study it was unclear whether it was prospective or retrospective (43).

Item 6-9. Methods-participants. Only one study (47) provided clear inclusion and exclusion criteria, while only the type of cancer of interest was described in the other studies. Patients were in eight studies identified based on inclusion in a registry. In one study it was unclear how participants had been identified for inclusion (43). One study included only patients who previously had tested negative for BRAF mutations (49), one study included those patents that previously had been tested with either DNA- or RNA-based NGS (11). Seven studies provided information on where and when eligible participants had been identified, and in two studies (11;50) this information was not complete. A majority of studies provided no information on whether participants formed a consecutive, random or convenience series. In one study the sample was described as 'unselected' (47), and in another study as 'consecutive' (49).

Item 10-13. Test methods. The intervention (Index) test, as well as the reference test, were described to some degree in all studies. However, not all three phases (pre-analytical, analytical, post-analytical) were clearly described, and it was unclear whether the detail provided was sufficient to allow replication. The authors typically did not use the terms index or reference tests when referring to the tests. No study provided

a rationale for choosing the reference standard /comparator test. For IHC (index test in all studies), a definition of what would constitute a positive test result was usually provided. These criteria however, varied across studies, and appeared to be more exploratory in some studies (11;45;50). As for the reference tests, FISH cut-off criteria for positivity were usually provided, while no information was provided on cut-off criteria/universal standards for PCR and NGS. In most studies no information was provided on whether clinical data and reference standard results were available to the performers/readers of the index test, and vice versa. Two studies (46;47) stated that the IHC analysis/interpretation was blinded among pathologists interpreting the test, but it was not clear whether they were blinded to the results of the reference standard.

Item 14-18. Analysis. In six studies there was no information on the methods for estimating or comparing measures of diagnostic accuracy (some studies had no statistics section). Only two studies provided a description of the methods used (43;45). Three studies provided some information on how different interpretation of IHC tests results among pathologists were handled (11;45;46), while the other five did not. One study (46) described how poor material that could not be analysed with NGS were handled. No studies included analyses of variability in diagnostic accuracy or provided any information on the intended sample size.

Item 19-22. Participants. None of the studies used a patient flow diagram. Five of the nine studies reported some baseline demographic and clinical characteristics of participants (45-49). One study provided information for NTRK positive samples only (44). Three studies reported the distribution of severity of disease in the study population (46;47;49). One study (45) described the time-period between tests, and in this study the tests were conducted simultaneously why no clinical intervention could have been administered in between tests.

Item 23-25. Test results. One study (43) cross-tabulated all cases (positive and negative cases with index test), and four studies (44;47-49) cross-tabulated cases positive with the index test against the results of the reference test. Five studies (11;43;45;47;48;50) provided a point estimate of diagnostic accuracy, but only two studies provided a measure of precision (i.e., 95% CI). No study reported on adverse events resulting from conducting the tests.

Item 26-27. Discussion. Five of the included studies (11;45-47;49) provided at least some discussion of study limitations and biases, while four studies did not (43;44;48;50). All studies provided some implications for practice, including the intended use and clinical role of the index test. Four studies proposed varying test algorithms, all with different starting points, for the detection of NTRK fusions in different type of cancer (11;47;49;50). These are described in a separate section below.

Item 28-30. Other information. None of the studies referred to a study registration, or to a published study protocol. Five of the studies had received non-commercial funding, and all authors of these studies declared no conflict of interest (43;45-47;49). Two studies provided no information on funding, but again all authors declared no conflicts of interest (44;50). One study was partly funded by commercial suppliers,

and many of the study authors had also received individual financial support from commercial suppliers (11). One study was conducted by commercial suppliers (48). The role of the funders was not reported in any of the original studies. See Appendix 10 Financial support, role of funders, and conflicts of interest.

Feasibility- advantages, and limitations of tests

IHC is a well-established, widely available, and relatively inexpensive analytical method (18;19;51). Also, FISH is relatively common, while RT-PCR is less so, and both techniques are relatively expensive. NGS is increasing in availability but is still quite expensive. The four methods can all have high sensitivity and specificity in certain settings (18). See footnotes Table 7.

Coverage- detection of fusion gene, partner, and protein expression

In contrast to the other methods- NGS can easily be multiplexed (52), which is a major advantage with this method. The four tests under study all cover the three NTRK fusions. However, FISH will need three probes- one for each fusion, RT-PCR may miss rare fusions, and DNA-NGS may not be capable of detecting fusions with large intronic regions e.g., NTRK2, and NTRK3 (18;51). IHC may sometimes result in false negative results, which have been suggested to mainly constitute NTRK3 gene fusions (18;53;55) False negatives have also been suggested to be due to sample preparation problems (e.g., fixation), highlighting the importance of the use of internal and external controls (19). IHC may further not be specific for NTRK gene fusion as it detects both wild-type and fusion proteins (give rise to false positive fusions), but in the absence of smooth muscle/neuronal differentiation the specificity is high (18;53-55). IHC, FISH, and RT-PCR provide no information on the fusion partner (18;51), while NGS can detect and characterise unknown partners. IHC allows correlation with histology, and can intrinsically confirm the protein expression, thus providing indirect evidence for NTRK gene fusions (18;51;55). The other tests provide direct evidence of a fusion (18;51).

Minimum read, quality requirements

Material requirements are lowest for IHC, and highest for NGS (18;19;53-55). FISH typically would require one probe for each of NTRK 1-3 (18;51;53). While IHC can work on poor materials, RT-PCR and NGS typically require good quality RNA/DNA for the analysis (51;53). Currently, liquid biopsies, is not a viable alternative for tumour-derived nucleic acids (and FFPE tissue) for the detection of NTRK gene fusions, as in most cases the existing DNA- or RNA-based NGS panels for liquid biopsy analysis, have limited coverage of NTRK gene fusions (18).

Cut-off criteria, reference standard, and clarity of test

There are no international cut-off criteria for positivity or scoring standard for IHC, and in addition IHC (51;54) has variable staining patterns for TRK expression, both in intensity and subcellular localisation (18).

For FISH however, there is a generally accepted scoring method for fusion gene detection, which include (i) counting the number of fluorescent signals in at least 50 randomly selected tumour cell nuclei (not overlapping); (ii) scoring preferably done by more than 1 observer; (iii) using a cut-off value of 10% or 15% break-apart (i.e. a gene fusion may be considered present if >10% or >15% of nuclei display 'split-apart signals'; and further that (iv) red and green signals should be separated by a distance greater than the size of two hybridisation probe signals. In addition, the sections used for the analysis should be 4 μm thick so artefactual split-apart signals can be avoided. There are as far as we know no universal cut-off criteria/standard for RT-PCR or NGS positivity.

Turnaround time (TAT)

The total TAT is the interval between when a test is requested to the time a treatment decision is made (56). While evidently some of the steps are the same for the different tests, the time requirement for the analysis, interpretation of results, and communicating the results to the treating physician in an accessible form, may differ. TAT is in addition dependent on whether the test is 'in-house', or if analyses are centralised to certain hospitals, or to facilities outside the hospital. IHC has a very short TAT, the TAT for FISH is slightly longer, while NGS can take as much as 1-3 weeks (18;19;53-55). Laboratories at Norwegian hospital trusts typically run NGS only one time/week (information from experts), although the use of NGS for cancer diagnostics is now rapidly increasing (29).

Facilities and expertise needed

High level of infrastructure and high-level bioinformatics capability is required for running NGS. Special expertise is also required for both FISH and RT-PCR, and special facilities for FISH (51). IHC can typically be applied in any clinical laboratory (18;54). Tools for interpretation of results and for communicating these to the treating physician, were not discussed in the included reviews.

Table 7. Characteristics of tests for the detection of NTRK gene fusions: summary of results from the narrative reviews and expert opinion papers (18;19;51-55)

	IHC	FISH	RT-PCR	RNA-NGS	DNA-NGS
Availability	Commonly available	Commonly available	Not widely available in Norway	Increasingly available	Increasingly available
Sensitivity, specificity	High[1,2]	High [3]	High	High	High[4]
Multiplexing	No	No	No	Yes	Yes
Detection of fusion gene	No	No	Yes	Yes	Yes
Detection of partner	No	No	No	Yes	Yes
Detection of expression	Yes	No	No	Yes	No
Minimum read (no of tumour cells/slides required)	Only a few cells/ a few unstained slides	At least 3 unstained slides (1 for each NTRK gene tested)	600-1000 cells	200 ng of RNA, (~10,000 cells)	250 ng of DNA (~50,000 cells)
Sample quality requirements	Works also on poor material	No information	Good quality mRNA	Good quality mRNA quality[5]	Good quality DNA[6]
Clarity of test	Variable staining pattern	No information	No information	No information	No information
Acceptable threshold (positive test)	No standardised scoring algorithm; no universal cut-off criteria	No information	No information	No information	No information
Coverage (detection of all fusion genes)	Yes	Yes, but one per probe	May miss some rare NTRK fusions	Yes	May not be capable of identifying fusions with large intronic regions (e.g., NTRK2 and NTRK3)[7]
Facilities and expertise required	Interpretation should take tumour histology into account (Solomon 2019)	Specific lab facilities and expertise for interpretation	Special expertise, and labour intensive (expert opinion)	High level of infrastructure and high-level bioinformatics capability	High level of infrastructure and high-level bioinformatics capability
Turnaround time	1-2 days	3-5 days	5–7 days	1-3 weeks	1-3 weeks
Potential role in test algorithm	Screening	Confirmation	Confirmation	Screening/ Confirmation	Screening/ confirmation

1 IHC: False negatives constitute mainly NTRK3 fusions (18); 2 Specificity is high in the absence of smooth muscle/neuronal differentiation (18). This because IHC may not be specific for NTRK gene fusion as it detects both wild-type and fusion proteins (18); Indication-specific specificity for NTRK gene fusion prediction not well characterized (unclear); Variable specificity according to tumour type (54;55); 3.FISH: High sensitivity but may have false negative results (51); 4 DNA-NGS: High sensitivity but with some caveats- may not detect fusions with large intronic regions; 5 Since RNA is more labile than DNA, good pre-analytics required to preserve RNA (52); 6 For FFPE tissue, sample age might affect DNA quality and sequencing read quality (52); 7 Detected rearrangements with DNA-NGS may not result in fusions, wherefore correlation with surgical pathology and predicted transcript (for sequencing) is a requirement (18).

Economic Evaluation

General

The health care sector, similarly, to society in general, is restricted by limited resources and budget constraints. In Norway, health service interventions are to be evaluated against three prioritization criteria: the benefit criterion (increased longevity and/or improved health-related quality of life), the resource criterion, and the severity criterion (absolute shortfall) (57). Norwegian policy documents indicate that the priority-setting criteria are to be evaluated together and weighted against each other. This is to be done by means of a health economic evaluation.

Health economic evaluations are important tools for decision makers facing questions on how to prioritize health technologies and maximize health benefits using limited resources. The basic aim of any economic evaluation is to identify, measure and compare health consequences and costs of the alternatives under consideration in an incremental analysis, one in which the differences in costs are compared with differences in health consequences.

Identifying the place of a molecular test within care pathways is crucial, not only to guide the selection of a relevant comparator, but also to guide the use of the companion drug and subsequent treatment pathways to be modelled. The exact place along the treatment pathway where testing occurs may change the cost-effectiveness of the intervention because of differences in the type of treatment subsequently received and the costs and outcomes arising from these (58).

In the published health economic evaluations of precision medicine, the cost of molecular testing and the accuracy of a test were reported as important factors, which influence the cost-effectiveness of targeted interventions (59;60). Ideally, diagnostic interventions should be supported by studies that follow patients from testing via treatment to final clinical outcome, so-called end-to-end studies (61). This combination of data will enable decision-makers to evaluate the overall cost-effectiveness of using a particular test-drug combination versus not using the drug at all or using the drug without the test (62). In other words, this will enable decision-makers to evaluate the predictive value of the test and the relevant biomarker, and also assess the health outcomes and economic consequences of using genomic test and the consecutive treatment in combination.

Key data needed for an economic evaluation of a molecular test include outcome data on the clinical effectiveness and utility of the technology, changes in health status as well as resource use and related costs of the affected patient population and the uptake of the test. Fundamentally, the challenges relating to the data requirements for the economic analysis of a pharmacogenetic intervention revolve around the availability and quality of existing data (58).

Method

Health economic evaluation of using molecular testing for NTRK gene fusions to determine eligibility for treatment with entrectinib or larotrectinib in patients with locally advanced or metastatic solid tumours, preferably, should be performed based on an integrated test-treatment model to follow the patient from the diagnostic test for the detection of NTRK gene fusions via treatment to clinical outcomes. Based on the results of our review, none of the included studies evaluated the tests' clinical utility. Moreover, the feedback from the suppliers of the diagnostic methods and the relevant pharmaceutical company supported the results of our review, i.e., there are no end-to-end studies available for the detection of NTRK gene fusions in patients with locally advanced or metastatic solid tumours.

Hence, in the absence of the relevant clinical utility data, we have in collaboration with the experts from the Norwegian regional health authorities, estimated the costs associated with each diagnostic method in Norway. The analyses were performed based on the micro-costing method. Micro-costing is a highly detailed health economic costing approach in which all the underlying resources required for an intervention or activity, such as equipment, consumables, and staff time are identified, and then unit costs are attached to this resource used to generate an overall cost (63).

Implementing precision medicine in healthcare is potentially a costly investment and it requires testing multiple patients to identify a specific group of responders to a targeted treatment. Currently, more and more multiple tests and multiple precision medicines for particular diseases become available (64). Almost in most of solid tumour cancers, e.g. in non-small cell lung cancer, a set of parallel tests are to be performed on a number of molecular biomarkers to decide between a range of precision medicines (64). Therefore, we also considered the multigene testing and testing samples from multiple patients in the estimation of the costs related to the diagnostic methods.

Furthermore, we calculated the costs per test associated with the relevant diagnostic methods based on the current Norwegian tariff rates for clinical laboratory services. A comparison between the estimated costs based on the micro-costing method and the costs calculated based on the tariff rates may assist the decision makers to make decision regarding the revision of the current tariff rates relevant for using genomic tests. We expressed relevant costs in 2021 Norwegian kroner (NOK).

Eligible population for detection of NTRK gene fusions

Due to lack of data on the incidence of patients with NTRK fusion solid cancers in Norway, we estimated the number of eligible people for testing of NTRK gene fusions based on two different scenarios:

In 2019, 34,979 new patients were diagnosed with cancer in Norway (65). Less than 1% of all cancers occurred in children. Cancers of the prostate, breast, lung, and colon were the most common cancers and accounted for 43% of all new cancers in Norway in 2019 (65). While NTRK fusions can be found at a lower incidence in more common solid tumours such as lung and gastrointestinal cancers, they are found in rare tumours such as secretory breast carcinoma, mammary analogue secretory carcinoma and infantile fibrosarcoma. There is great uncertainty in the number of patients eligible for NTRK testing due to the lack of certainty around NTRK fusion rates. If we assume that approximately an average of 35% of cancer patients have stage III/IV at diagnosis [1], or will experience recurrence, this will be result in about 11,140 patients (removed 3,158, number of blood cancer cases) who are eligible for NTRK testing each year in Norway.

Currently, there is no available registry of the number of NTRK fusion positive cancers in Norway, therefore, the number of patients eligible for NTRK testing might be estimated using yearly cancer deaths (for all ages) as a proxy. It was reported 11,049 deaths from cancer in Norway in 2018 (65). Since the proposed medicines only consider solid tumours, deaths due to blood cancers like leukaemia and multiple myeloma (about 965 patients) were removed from the estimation. The resulting number is 10,084 deaths due to solid tumours which can be illustrated the number of eligible people for testing to identify NTRK gene fusions.

Based on these estimations, we assume about 10,000-11,100 people are eligible for testing for NTRK fusions in Norway each year.

There is great uncertainty concerning how many of these patients had high frequency NTRK gene fusion tumours. However, based on the published rate incidence of these rare cancers, we estimated that each year about 520 patients with locally advanced or metastatic solid cancer can be eligible for testing to identify NTRK fusions with high frequencies (Table 8).

[1] There is a great variation regarding the risk for locally advanced or metastatic cancer associated with the different types of cancer. While the rate of stage III/IV is reported about 10% in melanoma (65), around 75% of NSCLS patients would have locally advanced or metastatic disease at the time of diagnosis (46).

Table 8. *Estimated number of patients eligible for detection of NTRK fusions with high frequencies*

	Number of patients	Comments
Adult cancers		
Mammary analogue secretory carcinoma (secretory salivary gland cancer)	1400	About 4% of all cancer patients (66)
Secretory breast carcinoma	6	Less than 0.15% of all cancers (67)
Paediatric cancers*		
Secretory breast carcinoma	3	Less than 1% of malignancy cases in childhood (68)
Infantile fibrosarcoma and other mesenchymal tumours	73	About 21% of all paediatric solid tumour (69)
Cellular and mixed congenital mesoblastic nephroma	1	3–10% of paediatric renal tumours (70), about 5% of all paediatric tumours are renal tumours (71)
Number of eligible patients with locally advanced or metastatic solid cancer for detection of NTRK fusions with high frequency	**520**	35% with stage III/IV (assumption)

** In 2019, less than 1% of all cancers occurred in children in Norway (65)*

Based on the estimation for the eligible number for NTRK testing in tumours with high frequencies, we have estimated that about 9,500-10,500 patients are eligible for testing to identify NTRK fusions with low frequencies in Norway each year.

Results

This section presents the costs associated with different diagnostic methods based on a micro-costing analysis and based on current Norwegian tariff rates for clinical laboratory services.

Costs associated with different diagnostic methods based on data of micro-costing analysis received from the Norwegian hospitals

To identify and measure the resource use and costs associated with the relevant diagnostic methods for the detection of NTRK fusions, we contacted four Norwegian regional health authorities. We received information about resources required for performing detection of NTRK fusions from three University hospitals: Stavanger University Hospital, St. Olav's University Hospital, and Oslo University Hospital. The estimated costs vary according to the approach used to estimate costs by different hospitals, however, all these estimates included direct costs such as consumables and supplies costs and the cost associated with staff time. The resource use and unit costs are presented in Table 9. It should be noted that the costs for NGS are estimated for a panel that can identify rearrangement in NTRK1,2,3, ROS1, ALK, and RET. For the other methods the costs were estimated for testing one biomarker, therefore the costs for testing several biomarkers with these tests (IHC, FISH) will be higher.

Due to data consistency, our analyses are based on information received from Stavanger University Hospital (72). Data sent from St. Olav's University Hospital and Oslo University Hospital are presented in Appendix 11.

Table 9. *Estimated cost for testing of one biomarker**

Diagnostic methods	IHC		FISH		NGS**	
	1 patient NOK	10 patients NOK	1 patient NOK	10 patients NOK	1 patient NOK***	10 patients NOK
Reagent costs	1200	1,200	1,500	1500	14,480	1,730
Personal costs	494 (Bioengineer, pathologist)	457	671 (Bioengineer, Molecular biologist, pathologist)	654	1,821 (Bioengineer, Molecular biologist, pathologist)	594
Sum	**1,694**	**1,657**	**2,171**	**2,154**	**16,301**	**2,325**

** Based on the feedback from the experts RT-PCR is not used to detect NTRK or other gene alterations (ALK/RET/ROS1) at pathology labs in Norway. We have therefore not included PCR in our cost-analysis.*

*** The panel can analyse 8 samples and 6 biomarkers simultaneously.*

**** The estimated costs are associated with testing one sample.*

The costs associated with IHC, and FISH analysis were estimated for the testing of one biomarker (Table 9). However, running a sequence of single-gene tests can be time-consuming and may require a relatively large tissue sample, which is not always available (73). In addition, the NTRK gene fusions detected in the relevant population may not be the only oncogenic driver mutation (11;48;74). Therefore, it is likely that some patients may become eligible for more than one targeted therapy at the same time. Furthermore, based on the experts' opinions, testing samples from several patients at once can save the use of resources and consequently the costs associated with the diagnostic methods.

Thus, in two different analyses, we have presented a set of parallel tests performed on a number of molecular biomarkers relevant for advanced NSCLC (testing 2 or 3 biomarkers at the same time), as an example.

> Scenario a: we presented the cost associated with using different diagnostic methods for testing ROS1 and NTRK in advanced NSCLC.
> Scenario b: we presented the cost associated with using different diagnostic methods for testing ROS1, NTRK, and ALK in advanced NSCLC.

In both scenarios, the costs are presented when we run the test for one patient or for ten patients. All estimated costs are based on data from Stavanger University Hospital (72). The results of these scenarios are presented in Table 10.

Table 10. *Estimated costs associated with using different diagnostic methods for testing 2 or 3 biomarkers for one or for 10 advanced patients with NSCLC*

Diagnostic methods	**Testing ROS1 and NTRK**		**Testing ROS1, NTRK and ALK**	
	1 patient NOK	**10 patients NOK**	**1 patient NOK**	**10 patients NOK**
IHC*	2,890	2,850	4,160	4060
FISH	7,020	4,160	10,150	5,860
RT-PCR**	3,450	1,490	3,870	1,310
NGS***	16,300	2,330	16,300	2,330

Source: (72)

** IHC positive results are required to be confirmed by the other methods.*

*** Based on the feedback from the experts, RT-PCR is not used to detect NTRK or other gene alterations (ALK/RET/ROS1) at pathology labs in Norway. We have therefore not included PCR in our cost-analysis.*

*** The costs are estimated for Oncomine Focus panel. The panel can analyse 6 biomarkers simultaneously including ALK, RET, NTRK 1, 2, 3 and ROS1*

As shown in Table 10, IHC is less costly for testing one biomarker in one patient. However, the costs will increase as more than one biomarker is going to be tested at the same time. Furthermore, IHC is just recommended as a pre-test for the detection of NTRK gene fusions. The positive results are required to be confirmed by the other relevant methods (18;26). Consequently, total costs can be increased for testing several biomarkers and samples. In addition, as previously shown in this report, IHC has lower sensitivity for detecting NTRK3 fusions as compared to detecting NTRK1 and NTRK2.

The results of the cost analyses showed that NGS can be one of the most affordable diagnostic methods. The method is considerably more expensive if only one patient (one sample) would be tested. However, the foremost advantage of NGS technologies is the massively parallel sequencing capability. It means sequencing of multiple targeted genomic regions from the multiple samples in the same run. As the results showed in Table 10, the cost associated with NGS testing will be significantly decreased (approximately NOK 2,000 per patient) when parallel tests are performed on several biomarkers from multiple patients.

The costs associated with using RT-PCR for testing several biomarkers and samples simultaneously are less than the costs related to the other tests. Although RT-PCR is a specific technique, it lacks somewhat in sensitivity and reliability. Rare fusion genes may be missed if the primer set for the multiplex PCR reaction does not cover the fusion gene in question, and quality mRNA may not be available from FFPE tissue (1). This technology is not widely available and requires special expertise. To our knowledge, none of the Norwegian hospitals uses RT-PCR for the detection of NTRK fusions and other gene alterations such as ROS1.

Costs associated with the detection of NTRK gene fusions

As presented earlier in this report, clinical diagnostic strategies, and the choice of the test for the detection of NTRK gene fusions will depend on the frequency and type of the NTRK gene fusion in a particular tumour. Therefore, the overall population relevant for NTRK fusions testing is divided into two sub-populations based on the NTRK gene fusion frequency, and the relevant testing costs were presented for these two sub-populations, i.e., solid tumours with a high frequency of NTRK gene fusions and solid tumours with a low frequency of NTRK gene fusions.

The European Society for Medical Oncology (ESMO) (18;26), based on the experts' recommendations has published the ESMO guidelines regarding the most reasonable strategy to adopt when screening for NTRK fusions in oncologic patients (18;26). According to our clinical experts, these recommendations are also relevant for implementation in Norway. Therefore, our costs estimations were based on the ESMO recommendations.

Adult and paediatric populations with histologically defined high frequency NTRK gene fusion tumours

Based on the recommendation from ESMO, the adult and paediatric cancers with a high frequency of NTRK gene fusions, require no prior tests and undergo either FISH, RT-PCR or RNA-NGS to identify the presence of NTRK gene fusions (18;26).

FISH can be quite effective at identifying the presence of the ETV6-NTRK3 fusion gene in the tumour with a high frequency of NTRK gene fusions (18). However, FISH cannot ascertain the 5´ partner nor whether the fusion results in a productive in-frame chimeric transcript (18). In addition, separate FISH assays would have to be run in parallel to detect different NTRK gene fusions. As shown in table 10, FISH will be more expensive than the other methods if it would be used for the detection of several biomarkers from multiple patients simultaneously. Furthermore, a multiplexed FISH becomes time-consuming and requires plenty of experience to interpret the results (18).

RT-PCR may also be used for the detection of NTRK gene fusion cancers with high frequencies. However, as previously mentioned, RT-PCR is not used to detect NTRK or other gene alterations (ALK/RET/ROS1) at pathology labs in Norway.

We have therefore included no prior tests in our cost analysis for the detection of NTRK gene fusions in solid tumours with a high frequency of NTRK gene fusions, and the patients undergo NGS to identify the presence of NTRK gene fusions. Results showed in Table 10, demonstrate that the cost associated with NGS testing will be significantly decreased (approximately NOK 2,000 per patient) when parallel tests are performed on several biomarkers from multiple patients. Assuming, in Norway, about 520 patients annually are eligible for testing to identify NTRK fusions with high frequencies, the costs are estimated to be about NOK 1,210,000.

Adult and paediatric populations with histologically defined low frequency NTRK gene fusion tumours

For patients, both adult and paediatric populations, with a low frequency of NTRK fusions tumours that are locally advanced or metastatic have been recommended that NGS targeted panel (preferably RNA-NGS) that reliably detects NTRK gene fusions would be ideal (18;26). Further, it has been recommended that if an NTRK gene fusion is identified, then the most exhaustive approach would be to include IHC to confirm protein expression of the detected NTRK fusions. Alternatively, a “two-step approach” could be considered, especially if NGS technology is not available, which includes IHC testing as a pre-test, and confirmation of any positivity detected with IHC by NGS (18;26). Our analysis is based on the second alternative.

IHC enables the detection of TRK overexpression as a surrogate for the presence of an NTRK gene fusion and provides a time-efficient technique that may be used for routine screening. However, it has been advised that IHC results must be followed by confirmatory testing using a molecular method to verify the presence of a fusion. This

is because overexpression of wild-type TRK proteins may also be detected (19). Patients with a positive IHC test will then undergo RNA-NGS analysis. The utilisation of NTRK fusions with low frequencies by using IHC testing as a pre-test and confirmation by using NGS is presented in Table 11.

***Table 11**. Estimated number of IHC positive NTRK fusions patients with low frequencies eligible for testing by NGS as a confirmatory test*

	Number of patients	Comments
Patients eligible for NTRK fusions testing with IHC	9,500-10,600 *	
IHC positive patients	(approx.) 30	87,9% sensitivity (11)
NTRK fusion-positive patients confirmed by NGS testing	**26**	RNA-NGS confirmation (100% sensitivity and specificity, assumption)

** For more information, please see Section "Eligible population for the detection of NTRK gene fusions".*
Approx.: approximately. The numbers are rounded.

Total costs associated with testing solid tumours with a low frequency of NTRK gene fusions by using IHC as a pre-test with RNA-NGS confirmation were estimated to be approximately NOK 16.1-18.0 million (Table 12). The costs per IHC testing and per NGS testing are based on the mean costs presented in Table 9. However, it should be mentioned that NTRK gene fusions with low frequencies are mostly detected in more common solid tumours, such as breast, colorectal, and lung cancer, which is often relevant for testing a number of other oncogenic alterations (11;48;74). Therefore, tissue removed during a biopsy can be tested for several molecular biomarkers simultaneously. This should be considered in the estimation of total costs for molecular testing of different biomarkers in this population.

***Table 12.** Estimated cost related to NTRK fusions with low frequencies by using IHC testing as pre-test with RNA-NGS confirmation*

	Number of eligible patients	Cost per test* (NOK)	Cost of testing NTRK fusions in eligible patients (NOK)
Cost of IHC testing	9,500-10,600	1,690	16,020,000-17,900,000
Cost of NGS testing	30	2,330	69,900
Total costs by using IHC as pre-test with NGS confirmation			**16,100,000-18,000,000**

**Source: (72). The numbers are rounded.*

Other relevant costs

Preparing the biopsy

According to our experts, the costs associated with preparing the biopsy are the same for all diagnostic methods. Therefore, the biopsy costs are not included in the comparison between diagnostic methods. Based on data we have received from St. Olav's University Hospital the cost per biopsy per patient is approximately NOK 270 (Table 13).

Table 13. *Biopsy costs per patient*

	Costs per biopsy (NOK)	Comments
Formaldehyde 20 ml	12.66	50 pieces per package
Biopsy forceps	163.50	20 pieces per package
Biopsy wraps	1.24	Bio-wraps (100 pieces per package)
Personnel costs: nurses	92.00	10 minutes per biopsy
Total cost per biopsy per patient	**269.40**	

Source: St. Olav's University hospital (personal communication)

The costs related to the infrastructure, quality assurance and maintenance

IHC and FISH

Based on the information from the experts, all hospitals in Norway can perform IHC. For FISH testing there is a need for a fluorescence microscope/scanner to be able to interpret the results in addition to the equipment for IHC. All university hospitals and some of the regional hospitals can perform FISH in Norway. However, the infrastructure costs for IHC and FISH are estimated to be around NOK 750,000 and NOK 2,750,000, respectively (Oslo University Hospital, personal communication).

Most pathology departments already have a maintenance plan and agreements on their IHC machines, thus there is no extra cost associated with the introduction of NTRK or ROS1 with IHC or FISH. External quality rounds (NordiQc, EMQN) are estimated to cost around NOK 5,000 per biomarker/per year.

NGS

Currently, about eleven pathology departments in Norway have equipment (six hospitals have Ion Torrent S5 and three hospitals MiSeq) to run RNA sequencing. Depending on the equipment capacity, it is estimated that the costs related to equipment and supplies investments are approximately 3 to 4 million NOK. The maintenance costs of NGS instruments are calculated to be about 150,000 per year.

Based on the information from the experts, the validation process takes a lot of time and different types of expertise (including bioengineers, pathologists, engineers) are involved in this process. At least one NGS kit is used for the validation process. The price of such a kit varies from NOK 30,000 to 80,000. The validation process is only performed once when the method is established. After that, external quality control system is used to check that all is still functioning adequately.

Costs associated with different diagnostic methods based on current tariff rates for clinical laboratory services

In Norway, outpatient laboratory services are financed on block grants, co-payments from the patients, and tariff rates for clinical laboratory services (75) while laboratory services for inpatients are financed on case-based payment (Diagnosis-related group, DRG).

In Table 14, we present the cost per investigation associated with different diagnostic methods used for the detection of NTRK gene fusions in Norway, based on the tariff rates. The average cost is calculated as the total of the tariff per investigation and the patient's co-payment, multiplied by two (76).

Table 14. *Average costs associated with using the relevant diagnostic methods* based on the Norwegian tariff rates*

Diagnostic methods	Costs per test (NOK)	Comments
NTRK IHC	721 (247+474)	705b:3-7 blocks, 705k: 4 or more analyses
NTRK FISH	449	701g: 1-3 probes
NTRK NGS**	14,507 (108+14,398)	701b: Organic extraction of DNA/RNA, 705s: DNA/RNA gene sequencing analysis package

Source: Lovdata poliklinikk-takster 2021 (75). The numbers are rounded.

**RT-PCR is not used for testing NTRK in Norwegian hospitals; therefore, it is not presented.*

***The tariff is used per analysis package, if both RNA and DNA sequencing are performed, the tariff can be used twice per patient. For testing NTRK, we included just the costs associated with RNA sequencing. The tariff is only used for outpatients or samples sent to the laboratory.*

A comparison between the estimated costs based on the micro-costing method and the costs estimated based on the current tariff rates indicated that the current tariff rates are generally insufficient to cover the costs of running IHC and FISH methods. Regarding NGS testing, the analyses showed that the relevant tariff rate can cover the costs of running the small NGS assay (for example, using Oncomine Focus assay gene fusion detection for 23 selected genes). Recently, the expanded gene panel used for the identification of the relevant cancer patients, mainly for the experimental treatment, which requires a thorough pathology assessment and a significantly more advanced data analysis, has received new tariff rates (Table 15) (75).

Table 15. *Average costs associated with using NGS for the identification of the relevant patients for experimental treatment based on the Norwegian tariff rates*

	Costs per test (NOK)	Comment
Reassessment of preparation for extended molecular analysis, selection of tissue area	353.50	Tariff rate 705t
Expanded gene panel analysis	14,398.48	Tariff rate 705u
Expanded data analysis and interpretation of sequencing data and reporting in interdisciplinary meetings	19,998	Tariff rate 705v

Source: Lovdata poliklinikk-takster 2021 (75)

Discussion

Summary of main results

We included nine original studies (6 comparisons) that compared the accuracy of IHC with one or more analytical techniques (i.e., FISH, RT-PCR, DNA-NGS, or RNA-NGS) for the detection of one or more NTRK fusions (NTRK1, NTRK2 and NTRK3) in patients with a variety of solid tumours (11;43-50). The number of included participants/samples ranged from 15 to 4,136 (Median: 60). Five narrative reviews (19;51-54), and two expert opinion papers (18;55) reported on advantages and limitations of the different tests.

Un-pooled results from six studies (involving 4 reference tests) suggest that the sensitivity of Pan TRK IHC for the detection of NTRK fusions may vary from around 40% to 100%, and specificity from around 20% to 100% across various types of solid tumours (11;43;47-50). The results of this review further suggest (i) Varying sensitivity of IHC for the three NTRK gene fusions (high for NTRK1, but lower for NTRK3 and NTRK2), (ii) Lower sensitivity for DNA-based as compared to RNA-based NGS for the detection of NTRK 2 and NTR3 gene fusions (11), (iii) Varying levels of consistency between cases positive with IHC and RNA-based NGS (range: 0 to 60%) (44;46;47), and (iv) Higher consistency between cases positive with FISH and RNA-based NGS, than between FISH and DNA-based NGS (47).

We did not identify any systematic reviews, or meta-analyses evaluating the accuracy of tests for the detection of NTRK gene fusions. Nor did we find any end-to-end studies with data on clinical utility, i.e., none of the included studies reported on outcomes of importance to patients (e.g., overall survival, quality of life), or how well the test(s) could predict the treatment effects (e.g., shrinking of the tumour or slowing down of the disease progress). No study reported on adverse events related to the testing.

Quality of included evidence and quality of reporting

According to the EGAPP quality assessment only one included study (11) provided relatively adequate evidence for the sensitivity and specificity of tests for detecting

NTRK fusions, while the other eight studies were judged to provide inadequate (low quality) evidence (see Appendix 8).

Three studies (43;45;46) verified only samples positive by one or more (index) tests with the reference standard (or a subset of samples with sufficient material), which rendered the rate of false negative samples unknown. In addition, most studies did not report the number of unevaluable samples, or described how indeterminate test results were managed, which also may have resulted in bias. Blinding of analysis and interpretation of the tests was unclear in all studies, which in the case of non-blinding may have affected especially the subjective interpretation of the IHC results. Some studies that were commercially funded (see Appendix 10), did not describe the role of the funder in the project, which may have introduced bias as it has been suggested to affect the results reporting in a direction beneficial for the funders.

The included studies also had much wanting in fulfilling the reporting standards according to the STARD reporting checklist (see Appendix 9). Many items crucial for adequate appreciation of the study results, and for replication of a study were missing. Future studies in this field would benefit greatly from the use of a reporting checklist.

Possible reasons for the differences in test accuracy

There are many possible reasons for the differences in test accuracy reported in this review, of which some are listed below.

Prevalence of fusion

- Varying prevalence of NTRK gene fusions across different types of solid tumours is a possible source of variability, as different biomarker prevalence may result in varying levels of test sensitivity and specificity (77).
- Cancer stage, as NTRK gene fusions may be more abundant in advanced or metastasised cancer, but disease stage was unknown in a majority of included studies.
- Pre-enriched samples may also affect the test accuracy, e.g., when all samples come from a major trial site with much higher NTRK prevalence than reported in the literature (11).

Test and analysis

- Storage of FFPE samples for more than 4-6 years may cause degradation of DNA, and in particular may this affect analysis with NGS (78).
- Different cut-offs for test positivity, different antibody clones, or test systems used may also have affected test accuracy.
- Non-blinding of those performing the test and interpreting the results may introduce bias.
- Varying time-period between index and reference tests, may introduce bias if other interventions are introduced to the patients in between tests.

Feasibility of tests

There are no perfect tests for the identification of NTRK gene fusions, i.e., that have low tissue requirements, can work on sub-optimal material, have clear cut-offs/standardised scoring, short TAT, good coverage (i.e., do not miss some fusions), can detect unknown fusions/partners, require little hands-on-time, do not require ex-tensive expertise and infrastructure to run and interpret the analyses, has the capacity to analyse multiple genes simultaneously, and all at a low cost. Some of the included original studies (47;49;50) suggested to use IHC for initial screening, but with confirmation of IHC positive cases with other methods (e.g., RNA-NGS or RT-PCR), while one study suggested to use RNA-NGS directly, and when needed to confirm protein expression with IHC . IHC (11) is not recommended to be used for screening of high frequency tumours, due to its lower sensitivity for NTRK3 fusions which are common (usually ETV6-NTRK3) in the more rare high frequency NTRK tumours e.g., in infantile fibrosarcomas and secretory carcinomas of the breast and salivary glands which typically have >80% NTRK gene fusions (77).

It should be noted that the recommendations from ESMO published in 2019 (18) regarding NTRK testing in people with solid tumours, which also propose a test algorithm, are based on evidence from un-pooled results from single studies (and expert opinions) that is neither quality assessed, nor is the certainty of the evidence graded.

Short TAT, enabling timely administration of targeted drug treatment without unnecessary delay, potentially play an important role especially for patients with aggressive fast-growing tumours. There are as far as we know no studies available that have compared the timeliness of treatment administration related to different tests. Results from a modelling study, conducted in the US and focussing only on patients with NSCLC, suggest that NGS analyses would save time compared to sequential single-gene analyses involving a number of different biomarkers in NSCLC (79). However, as there appear to be relatively large differences in TAT between different NGS systems, and the number of biomarkers relevant to test for may differ across solid tumours, the time that potentially can be saved, as compared to sequential single-gene tests, must depend heavily on the type of NGS system used, and the number of relevant biomarkers to be tested. The organization of test services also plays an important role for the TAT and thus for the timeliness of the treatment decisions, e.g., whether the laboratories in question run NGS every day, or once a week, which at present is the case at Norwegian hospital trusts. The development of a testing algorithm for the detection of NTRK fusions will depend on many factors: histology, accessibility of testing modalities, and economic considerations.

Patient preferences

In our previous report (1), we discussed patients' preferences related to pharmacogenomic testing, which appear to depend on several factors, of which some are related to the test per se (55). In summary, there are three test-related factors that appear to be of particular importance to patients: (i) the invasiveness of the test, (ii) the sensitivity and specificity of the test, and (iii) the prevalence of the biomarker of interest (55). So how do these factors relate to the tests under study?

(i) Invasiveness of test

Even though NGS require considerably more input material, than both IHC and FISH it has, due to its capacity to analyse numerous genes simultaneously, been suggested to have a potential to conserve material and thereby reducing the risk of repeat (invasive) biopsies, as compared to sequential testing, especially in lung cancers with limited size tissue biopsies (52). However, none of the studies included in this review addressed this issue. The use of liquid biopsies would reduce the need for repeat tissue biopsies, but at present liquid samples are not used for the detection of NTRK gene fusions, as in most cases the existing DNA- or RNA-based NGS panels for liquid biopsy analysis, have limited coverage of NTRK gene fusions (54). There is ongoing research and method development related to the use of liquid biopsies in pharmacogenomic testing, which if moved into routine use, would relieve patients from the pain and unpleasantness of repeat tissue biopsies (80).

(ii) Sensitivity and specificity of the test

The sensitivity and specificity of a test is of importance to the patient as high rate of false negative tests could result in a missed treatment opportunity, and a high rate of false positive tests could increase the risk for having to endure an unnecessary and non-effective drug treatment, possible with adverse effects. The sensitivity and specificity of tests is probably not an issue if IHC is used to screen for NTRK gene fusions in low-frequency tumours (that typically do not contain NTRK3 gene fusions that IHC may miss detecting) and another test (e.g., RT-PCR or NGS) is used for confirmation. For high frequency tumours (with NTRK3 fusions) RNA-NGS should be used directly, without pre-screening with IHC.

(iii) Prevalence of the gene fusion/alteration

NTRK gene fusions are rare in most solid tumour (<1%). There may however be other targetable gene alterations of relevance for patients with solid tumours. If other relevant biomarkers (with available targeted treatment) are simultaneously tested for using multi-gene panels, the potential for finding targetable gene mutations increases, which may be a motivator for the patient to be tested. However, as there are more than 100 various types of cancer, we do not currently know the number of biomarkers relevant to test for (with available targeted treatment) in each one of these tumours, we only know that the number of targetable mutations is steadily increasing.

Strengths and limitations

The HTA

This is the second assessment of the accuracy of tests for the identification of targetable gene alterations and fusions in patients with cancer conducted at NIPH. The report may be seen as a second pilot of our new assessment framework for pharmacogenomic tests, in which we this time around also assessed the quality of included evidence with the EGAPP-tool, as well as the quality of reporting using the STARD checklist. We hope this report will be of further help in detailing future commissions, so as to ensure that our reports are of help to decision makers.

We developed a robust search strategy and conducted a comprehensive search for primary studies (and systematic and non-systematic reviews). We conducted duplicate screening, data extraction and quality assessment, all to minimize bias and reducing the risk of missing important evidence.

In our recent publication of tests for the detection of ROS1 gene alterations (1), we did not identify, any reviews concerned with the ethical, social or legal impact (ELSI) of pharmacogenomic testing, (e.g. confidentiality issues, disclosure of genomic test results, ownership of data, panel testing for mutations with no available treatment, and the costs of scaling up of pharmacogenomic testing and targeted therapies) (51). We did not update the search for this report as we believe that ELSI, due to its complexity, need to be addressed in a separate publication. We did not update the search for studies of patient preferences related to pharmacogenomic testing, and the discussion concerning patient preferences rely on the findings from our previous publication, which may be seen as a limitation with this report.

Results on test accuracy are reported narratively as meta-analysis was not feasible due to heterogeneity across included studies in terms of type of cancer, prevalence of NTRK gene fusions, type of outcome measures reported (i.e., sensitivity and specificity, AUROC, concordance, and consistency), and comparator (reference) test used. In addition, few studies reported a measure of dispersion.

This review did not focus on studies comparing the accuracy or feasibility of different NGS-systems (or different types of FISH or PCR tests), and therefore such publications were not included in this HTA. An exception to this was the study by Salomon et al (11), which in addition to reporting on the accuracy of IHC (as compared to RNA-based NGS) for the detection of NTRK fusions, also provided results for a comparison of DNA- and RNA-based NGS.

We included a handful of narrative reviews, and a couple of expert opinion papers to address the advantages and limitations of the different tests. While systematic reviews and meta-analyses are considered the highest level of evidence, narrative reviews, and expert opinion papers, are generally considered to be low-level evidence (78). Normally we would not include this type of publications in an HTA, but as they appear to

provide valuable information on the feasibility of tests, we chose to include these publications to demonstrate the construct of the new framework we are piloting in this HTA.

As mentioned earlier, we did not find any end-to-end studies, and thus have no evidence concerning outcomes of importance to patients (e.g., overall survival, quality of life), or on how well the tests can predict the effectiveness of treatment (e.g., shrinking of the tumour, or slowing down the disease process). Nor can we determine whether actually taking a test, as compared to not taking a test, would make a difference for overall survival.

The included evidence

Eight of the nine original studies included in this HTA, provided inadequate evidence according to the EGAPP quality assessment (one exception was Salomon 2020), and the reporting was overall poor. See section Quality of evidence and quality of reporting section for more detail.

The studies were heterogenous in terms of type of cancer, prevalence of biomarker, outcome measures reported, comparator (reference) test used, cut-off criteria for positivity, age of samples (duration of storage), to mention some of the many factors that may have affected the results of the included studies.

In three of the nine studies (43;45;46) only positive cases with index test (or sample with sufficient tissue) were verified with the reference test, which hampered any attempt to calculate the sensitivity and specificity of the index test. However, the number of false negative cases should be very low as these cancer types are not known to include NTRK3 gene fusions, for which IHC have lower sensitivity for (77).

Mostly single studies reported results for test accuracy for the various cancer types included, why it was not feasible to report the results separately for each single type of solid tumour. For five cancer types (i.e., CRC, lung cancer/NSCLC, thyroid, STS, and salivary gland carcinomas) more than one study provided test accuracy data, but they reported different accuracy related outcomes and varied greatly in prevalence of NTRK fusions. In addition, while there are more than 100 different types of cancer, this review has only covered some of them.

The included studies typically provided very little information on the characteristics of participating patients other than the type of cancer included. Therefore, in most cases, we do not know the proportion of patients in the included studies who had advanced or metastatic disease, which was the actual group of interest in this report. Nor do we know if any, or all, of the participants had previously been, or were being, treated with Trk inhibitors.

Other limitations with the included evidence are addressed in the section on quality of included studies, and quality of reporting.

Economic evaluation

Several testing strategies for the detection of NTRK gene fusions in locally advanced or metastatic solid tumours exist, however there is still a lack of evidence on the comparative economic implications of using these strategies in Norwegian clinical practice. We assessed the costs associated with the relevant diagnostic methods based on the data received from the Norwegian University Hospitals, the laboratories of molecular biology.

Our assessment showed that the costs with NGS approach are likely to be higher than the other diagnostic methods if it is used for one patient only, due to the higher reagent costs of the NGS test. However, as NGS technology allows massively parallel sequencing and testing samples from several patients at once, the cost associated with NGS testing will significantly be decreased when parallel tests are performed on several biomarkers from multiple patients. In addition, the need for tissue preservation and the burden and comorbidity of repeat biopsies is likely to decrease.

The results showed that the cost associated with using NGS was around NOK 16,000 per sample. However, if several samples and biomarkers are tested with NGS at the same time, e.g., NTRK, ROS1, RET and ALK in NSCLC patients, the costs will be reduced to approximately NOK 2,000 per patient.

We have estimated the costs associated with the detection of NTRK gene fusions based on the frequency of the NTRK gen fusions in different types of solid tumours. Assuming about 520 patients annually are eligible for testing to identify NTRK fusions with high frequencies, the costs were estimated to be about NOK 1.2 million. In tumours with a low frequency of NTRK gene fusions, the costs associated with NTRK testing by using IHC as a pre-test with RNA-NGS confirmation were estimated to be approximately NOK 16.1–18.0 million. The costs did not include overhead, capital, and other infrastructure costs.

It is notable that the costs estimations are highly dependent on the number of people that will be tested for NTRK gene fusions. In addition, almost in the most solid tumours a set of parallel tests are to be performed on several molecular biomarkers to decide between a range of precision medicines (79). Comprehensive fusion testing (for all major sarcoma fusions) is increasingly being carried out as a first-line test in sarcomas. Inclusion of NTRK testing in comprehensive sarcoma fusion test panel has been therefore recommended (55). Specially, in certain indications, which we would anticipate increasingly in the future, NGS will be routinely requested as part of the diagnostic workup (e.g., NSCLC) (52). Therefore, it is intrinsically inefficient if one is solely screening for an alteration with very low prevalence, such as NTRK fusions (19). In the estimation of the costs related to testing for NTRK gene fusions with NGS, we should therefore consider that the multigene would be tested at the same time.

Based on the data from the Norwegian pathology departments, the capital and infrastructure costs (including overhead costs) are higher for NGS than the other diagnostic methods. It has been estimated that the costs related to NGS equipment and supplies investments is approximately NOK 3-4 million. Further, the costs associated to maintenance the NGS instruments is more expensive (around NOK 30,000- 80,000) than the other methods. The validation process for any of the techniques is challenging as there are so few positive cases reported in Norway. A validation with 5-10 positive cases would be preferable but this is in practice very difficult. As such, the validation process at local hospital is dependent on positive cases/controls from other countries/companies.

A comparison between the estimated costs based on the micro-costing method and the costs estimated based on the current tariff rates indicated that the current tariff rates are generally insufficient to cover the costs of running IHC and FISH methods. For NGS testing, the analyses showed that the current tariff rate can cover the costs of running the small NGS assay. In addition, the expanded gene panel used for the identification of the relevant cancer patients mainly for the experimental treatment has recently received new tariff rates which can be used for the larger NGS panels.

Multiple labs also reported differences in reimbursement for internal versus external testing (inpatient versus outpatient) which may led to apparently unintended consequences, such as clinicians triaging groups of patients to prioritize for external testing when assays were not available in-house, or electing to perform diagnostics in an outpatient rather than inpatient setting if possible (80).

Currently, NTRK testing is not routinely done for all solid tumours in Norway. However, NGS is used at some Norwegian University Hospitals for the identification of NTRK gene fusions (personal communication).

About eleven Norwegian hospitals have invested in NGS technology and some of them have already access to the equipment of two commercial suppliers of NGS (Thermo Fisher and Illumina). The price sent from Illumina (TruSight™ Oncology 500, covers 523 cancer-relevant genes)[1] and Thermo Fisher (Oncomine™ Focus Assay, gene fusion detection for 23 selected genes), is approximately [redacted] and [redacted] per sample, respectively.

In addition, we have received the information from three other NGS suppliers (Roche, Caris' Life Science and Archer). For using the technology of these three suppliers, the patients' tumour samples should be sent abroad to the external laboratories. After

[1] We did not receive cost information for the smaller gene panels from Illumina (e.g., Illumina AmpliSeq 52 genes).

preparation and conducting the test, a clinical and biological report will be sent to the local hospitals. If eligible, they will return of remaining parts of the tumour block. One can discuss that this approach might result in some cost-saving due to the reduction the costs related to the work performed by local experts at the pathology department. However, it is important to mention that precision medicine is an interdisciplinary field that requires multidisciplinary collaboration among different field of expertise including pathology, oncology, and the laboratory. In addition, the legal and ethical consequences of this approach should be considered and assessed. The list price sent from these three suppliers (Roche, Caris' Life Science and Archer) is approximately between NOK 25,000-37,000. Roche offers a net price [redacted] excluding value added tax (FoundationOne® CDx, covers 324 genes). The price includes all the relevant procedures from pick up and transport of tumour sample to deliver the clinical report and return of remaining parts of the tumour block if eligible. Archer and Caris are also opened to discuss the price of RNA-sequencing analysis.

Although, we have tried to conduct our analysis based on the best available data, lack of cost data comparing different diagnostic methods was the most important limitation of this economic analysis. We contacted all Norwegian regional health authorities. We received data from three hospitals based on the relative different approaches. The data related to different diagnostic methods for NTRK gene fusions testing were presented in different scenarios by Stavanger University Hospital, and due to data consistency, our analyses are based on the data received from this hospital.

NTRK gene fusions are rare, less than 1% of solid tumours, however the number of patients eligible for treatment with tyrosine kinase inhibitors, like larotrectinib or entrectinib are highly dependent on the number that will be tested for NTRK gene fusions. There is great uncertainty regarding the number of eligible people for testing to identify NTRK gene fusions in Norway. This is mainly due to uncertainty around NTRK fusions rates. In the absence of a registry of the number of NTRK fusion positive cancers, we have estimated the number of eligible people for NTRK gene fusion testing based on the two different scenarios and some assumptions.

NTRK fusions can be found at a higher frequency in most rare tumours such as secretory breast carcinoma, mammary analogue secretory carcinoma and infantile fibrosarcoma. We could not find the incidence rates for these rare cancers in Norway. Therefore, our estimations were based on the published international data. However, the incidence rates presented in the literature were mostly based on a small number of cases, and there were some variations in the rates reported by different studies.

The costs associated with testing for NTRK gene fusions in low frequency tumours with IHC as prior test, followed by confirmation NTRK NGS testing were estimated based on the 100% test accuracy for RNA-NGS testing. Although it is reported that RNA-NGS has high accuracy for the detection of NTRK gene fusions, it is reasonable to assume that 100% sensitivity and specificity is only true in ideal circumstances. The

size of the population eligible for NTRK testing with NGS (as a confirmation test) is also dependent in NTRK IHC sensitivity and specificity in Norwegian practice.

We have not considered the additional re-biopsy costs in our analyses.

Challenges

Recently, the number of economic evaluations on precision medicine has noticeably increased. However, an important number of these analyses focus purely on the assessment of the actual therapeutic treatment, failing to include the impact that the actual tests have on the overall economic value of the test–drug combination (81). Even, when both testing and therapeutic decisions are taken into consideration, the weight of the accuracy of the companion tests on the overall results is rarely explored (81).

In principle, the healthcare system benefits from the availability of companion diagnostics that accurately identify responders, reduce the number needed to treat, and thereby improve the efficient use of resource. Therefore, the consequences of implementation of test-and-treat interventions and system integration challenges should be considered by the reimbursement authorities. In addition to that a good biomarker test should measure the biomarker with a high degree of accuracy and demonstrate analytical validity, the biomarker test should optimally demonstrate clinical utility, meaning that it improves patient's outcomes compare to a no-testing approach. This requires establishing the relationship between the test results and the consecutive treatment, and outcomes. Generating evidence to support the economic case of a precision medicine in practice, however, can be a challenge. Manufacturers, analysts, and funders of research may improve their research and development activities by considering the evidenced required by later-stage decision-makers at an earlier time period in the process of evidence generation.

The basic principles of cost-effectiveness should be applied to biomarkers. However, cost-effectiveness estimates for recent pharmaceutical-diagnostic combinations have been highly variable among major HTA markets, suggesting that methods for incorporating test information into economic evaluations are inconsistent. Key issues include gaps in the evidence supporting clinical utility and cost effectiveness of diagnostics (82).

Usually, information on treatment patterns and on the costs and outcomes relating to using diagnostic methods, is the most common limitation of the published economic evaluation of precision medicine interventions, especially data about false-positive and false-negative test results (82). The higher specificity rates of the diagnostic method will help to reduce the potential treatment of 'false positives' and consequently engaging in high spending for a proportion of the patient population for which the targeted therapy would not be effective or for which it could even produce some harm. San Miguel (81) have shown that in the field of targeted therapy in oncology,

test accuracy becomes even more crucial given high effectiveness but also high prices of some of these therapies. The specificity of a diagnostic method becomes even more important if very small population subsets are to be identified using the biomarker, as is the case for somatic mutations that are often present in no more than 1 to 2% of a specific tumour type. San Miguel (81) concluded that the importance of test specificity is twofold; for the patient it is crucial to receive the correct targeted treatment; for the society the use of (often expensive) targeted treatment in patients that do not benefit from it because the marker was not tested accurately, is a waste of money.

Furthermore, the health economic evaluation can even become more complex if different tests are combined or sequentially used. This potential complexity can be handled by explicitly showing how these tests are going to be used in practice and then working with the combined sensitivities and specificities of the tests (62). Moreover, the tests available and the test sequence employed in the clinical study, may differ in their ability to accurately select patients who will likely benefit from target therapy (83).

In addition to the issues related to properly assessing the health economic consequences of test-drug mentioned above, there are regulatory barriers to the development and adoption of precision medicine. First, the regulation of marketing approval is insufficiently harmonized. It varies across countries and is different for drugs and diagnostic tests. In the United States, marketing approval for drugs and diagnostics is done by the FDA. The joint approval process performed by a single agency ensures scientific knowledge-sharing and provides an effective way to approve precision medicines. However, in Europe, no single European agency regulates both medicines and tests. The European Medicines Agency (EMA) regulates the marketing approval of drugs, whereas it is each European Union (EU) member state's Notified Body that monitors the performance standards of diagnostic test (84).

Furthermore, the requirements for marketing approval of tests are still relatively lenient. In Europe, the test manufacturer is currently required to demonstrate the clinical validity (predictive capability) but not the clinical utility (effect on clinical outcomes) of the test. Another important challenge related to assess the cost-effectiveness of diagnostic methods is that, currently, the manufacturer of a new test does not need to demonstrate its effectiveness if a similar test already exists. Moreover, laboratory developed tests, that is, tests performed within a single laboratory or hospital (not commercialized) do not require a full regulatory review (84). Therefore, there is a lack of standardized evidence of the performance of biomarker tests in terms of their impact on health outcomes. This results in uncertainty for health authorities who make decisions on pricing and reimbursement based on the value of treatment produced by the biomarker test.

However, in 2017 the EU parliament and council agreed on a new set of regulations on in vitro diagnostics (85). Based on these new regulations, companion diagnostics

will need to meet stricter performance requirements, including clinical evidence and there will be a link between the assessment of a diagnostics by a notified body and the corresponding medical product by a medicine regulatory authority. It means that if a companion diagnostic is necessary to identify whether a patient is likely to benefit from a corresponding medical product, the evidence regarding its impact on patient outcome, i.e., clinical utility, will be carefully considered by the medicine authorities, in determining the benefit/risk of the medical product. Consequently, these evidence on clinical utility of companion diagnostics can be used in evaluation of health economic consequences of the test and the consecutive treatment. The regulations will become full effective in 2022 (85).

Conclusion

The results of this HTA were based on mostly inadequate evidence from nine studies on the accuracy of tests (IHC, FISH, RT-PCR, RNA-NGS and DNA-NGS) for the detection of NTRK gene fusions in solid tumours.

Un-pooled results including five tests and six test comparisons suggested varying test accuracy mostly for single gene testing (e.g., IHC), across different types of solid tumours and NTRK fusions. The results further suggest higher sensitivity of RNA-NGS than DNA-NGS in detecting NTRK fusions, especially for fusions with large intronic regions (NTRK2, and NTRK3).

While there are advantages and limitations for all tests, sequential single gene testing may be unfeasible, especially when the number of actionable biomarkers relevant for testing appear to be increasing.

There is a general agreement in included studies that IHC samples that are NTRK fusion-positive, need confirmation with other molecular methods (e.g., RT-PCR or RNA-NGS), due to a tendency of IHC for false positive staining. It is notable that IHC is not recommended to be used for screening of high frequency tumours that typically involve NTRK3 fusions, due to its lower sensitivity for these fusions. Overall, the development of a testing algorithm for the detection of NTRK fusions depends on accessibility of testing modalities, economic considerations, histology (i.e., based on a low and high frequency of NTRK gene fusions), and turnaround time.

We did not identify any systematic reviews, or meta-analyses evaluating the accuracy of tests for the detection of NTRK gene fusions. Nor did we find any end-to-end studies with data on clinical utility, i.e., none of the included studies reported on outcomes of importance to patients, or how well the test(s) could predict the treatment effects.

The cost associated with NGS testing will decrease significantly when parallel tests are to be performed on several biomarkers from multiple patients (using gene panels). However, at present, the capital and infrastructure as well as maintenance costs are higher for NGS than the other diagnostic methods.

The costs related to testing the solid tumours with a high frequencies og NTRK gene fusions were estimated to be about NOK 1.2 million. In tumours with a low frequency of NTRK gene fusions, the costs associated with NTRK fusion testing by using IHC as a pre-test with RNA-NGS confirmation were estimated to be approximately NOK 16.1-18.0 million. The costs estimations are highly dependent on the number of people that will be tested for NTRK gene fusions. However, the number of eligible people for testing in Norway is still uncertain. In addition, it is expected that screening for NTRK

fusions in common cancers with very low prevalence would be included in comprehensive fusion test panels.

The current tariff rates are generally insufficient to cover the costs of running IHC and FISH methods. The reimbursement rate for NGS testing can cover the costs of running the small NGS assays. In addition, new tariff rates have recently been defined for the expanded gene panels that mainly used for the experimental treatment.

Biomarker tests should optimally demonstrate clinical utility, meaning that it improves patient's outcomes compared to a no-testing approach. The consequences of implementation test-and-treat interventions and system integration challenges should be considered by the reimbursement authorities.

Future research should focus on conducting large cohort studies with well-defined patient populations, that follows the patients from testing (or no testing), through treatment and final outcomes. Further, studies should use robust and replicable methods, and follow reporting standards for diagnostic test accuracy reviews for improved clarity.

References

1. Flodgren G, Hamidi V. Tests for detection of ROS1 gene alterations in people with non-small cell lung cancer (NSCLC). A health technology assessment Norwegian Insitutute of Public Health; 2021.
2. Awad K, Dalby M, Cree I, Challoner B, Ghosh S, Thurston D. The precision medicine approach to cancer therapy: part 1 — solid tumours. Pharmaceutival Journal 2019.
3. Robinson D, Wu Y, Lonigro R, et al. Integrative clinical genomics of metastatic cancer. Nature 2017;548(7667):297–303.
4. Tsimberidou AM, Hong DS, Wheler JJ, Falchook GS, Janku F, Naing A, et al. Long-term overall survival and prognostic score predicting survival: the IMPACT study in precision medicine. J Hematol Oncol 2019;12(1):145.
5. Keeling P, Cark J, Finucance S. Challenges in the clinical implementation of precision medicine companion diagnostics. Taylor & Francis Group 2020;20:593-9.
6. Burke W, Atkins D, Gwinn M, Guttmacher A, Haddow J, et al. Genetic test evaluation: information needs of clinicians, policy makers, and the public. Am J Epidemiol 2015;156(4):311-8.
7. Phillips KA, Deverka PA, Hooker GW, Douglas MP. Genetic Test Availability And Spending: Where Are We Now? Where Are We Going? Health Aff (Millwood) 2018;37(5):710-6.
8. IQN Path, ECPC, EFPIA. Improving cancer care through broader access to quality biomarker testing- Policy recommendations. EFPIA; 2021.
9. Lange AM, Lo HW. Inhibiting TRK Proteins in Clinical Cancer Therapy. Cancers (Basel) 2018;10(4):04.
10. Vaishnavi A, Le AT, Doebele RC. TRKing down an old oncogene in a new era of targeted therapy. Cancer Discov 2015;5(1):25-34.
11. Solomon JP, Linkov I, Rosado A, Mullaney K, Rosen EY, Frosina D, et al. NTRK fusion detection across multiple assays and 33,997 cases: diagnostic implications and pitfalls. Mod Pathol 2020;33(1):38-46.
12. American Cancer Society. What is Cancer? The American Cancer Society medical information; 2020. Tilgjengelig fra: https://www.cancer.org/content/dam/CRC/PDF/Public/6041.00.pdf
13. Gavhane YN, Shete AS, Bhagat AK, Shinde VR, Bhong KK, Khairnar GA, et al. Solid Tumors: Facts, Challenges and Solutions. International Journal of Pharma Sciences and Research (IJPSR) 2011;2(1):1-12.
14. Drilon A. TRK inhibitors in TRK fusion-positive cancers. Ann Oncol 2019;30 Suppl 8:viii23-viii30.
15. Ricciuti B, Genova C, Crino L, Libra M, Leonardi GC. Antitumor activity of larotrectinib in tumors harboring NTRK gene fusions: a short review on the current evidence. Onco Targets Ther 2019;12:3171-9.
16. Westphalen CB, Preinfalk A, Kruger S, Haas M, Renz BW, Riener MO, et al. Neurotrophic tropomyosin receptor kinase (NTRK) and nerve growth factor (NGF) are not expressed in Caucasian patients with biliary tract cancers: pooled data from three independent cohorts. Clinical & Translational Oncology: Official

Publication of the Federation of Spanish Oncology Societes & of the National Cancer Institute of Mexico 2019;21(8):1108-11.
17. Forsythe A, Zhang W, Phillip Strauss U, Fellous M, Korei M, Keating K. A systematic review and meta-analysis of neurotrophic tyrosine receptor kinase gene fusion frequencies in solid tumors. Ther Adv Med Oncol 2020;12:1758835920975613.
18. Marchio C, Scaltriti M, Ladanyi M, Iafrate AJ, Bibeau F, Dietel M, et al. ESMO recommendations on the standard methods to detect NTRK fusions in daily practice and clinical research. Ann Oncol 2019;30(9):1417-27.
19. Penault-Llorca F, Rudzinski ER, Sepulveda AR. Testing algorithm for identification of patients with TRK fusion cancer. J Clin Pathol 2019;72(7):460-7.
20. FDA approves larotrectinib for solid tumors with NTRK gene fusions [press release]. Food and Drug Administration2018.
21. European Medicine Agency. Vitrakvi. European Medicine Agency; 2019.
22. Marcus L, Donoghue M, Aungst S, Myers CE, Helms WS, Shen G, et al. FDA Approval Summary: Entrectinib for the Treatment of NTRK gene Fusion Solid Tumors. Clin Cancer Res 2021;27(4):928-32.
23. European Medicine Agency. Rozlytrek, entrectinib. European Medicine Agency,; 2020.
24. Delgado J, Pean E, Melchiorri D, Migali C, Josephson F, Enzmann H, et al. The European Medicines Agency review of entrectinib for the treatment of adult or paediatric patients with solid tumours who have a neurotrophic tyrosine receptor kinase gene fusions and adult patients with non-small-cell lung cancer harbouring ROS1 rearrangements. ESMO open 2021;6(2):100087.
25. Metodevarsel. Rozlytrek. Mednytt: 2019.
26. OncologyPro.ESMO. Algorithm for NTRK gene fusion testing. ESMO: 2021. Tilgjengelig fra: https://oncologypro.esmo.org/oncology-in-practice/anti-cancer-agents-and-biological-therapy/targeting-ntrk-gene-fusions/algorithm-for-ntrk-gene-fusion-testing
27. Oslo Universitetssykehus. NGS diagnostikk innen kreft (somatiske forendringer). Oslo, Norway: 2020.
28. FDA Approves Companion Diagnostic to identify NTRK fusions in solid tumors for Vitrakvi [press release]. FDA2020.
29. FoundationOne®CDx Receives FDA-Approval as a Companion Diagnostic for VITRAKVI®(larotrectinib) to Identify Patients with NTRK Fusions Across All Solid Tumors [press release]. Foundation Medicine, Inc.2020.
30. European Medicine Agency. Rozlytrek, CHMP assessement report. European Medicine Agency; 2020.
31. List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools) [press release]. Food and Drug Administration2021.
32. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med 2009;11(1):3-14.
33. Veenstra DL, Piper M, Haddow JE, Pauker SG, Klein R, Richards CS, et al. Improving the efficiency and relevance of evidence-based recommendations in the era of whole-genome sequencing: an EGAPP methods update. Genetics in Medicine 2013;15(1):14-24.
34. Pitini E, D'Andrea E, De Vito C, Rosso A, Unim B, Marzuillo C, et al. A proposal of a new evaluation framework towards implementation of genetic tests. PLoS One 2019;14(8):e0219755-e.
35. Beynon R LM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et a.. . Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. Cochrane Database Syst Rev 2013;9.
36. Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions. AMIA Annu Symp Proc 2006;2006:359-63.
37. The EndNote Team. EndNote. EndNote X9 utg. Philadelphia, PA: Clarivate

2013.
38. Ouzzani M, Hammady, H, Fedorowicz, Z, and Elmagarmid. A. Rayyan — a web and mobile app for systematic reviews. Systematic Reviews 2016;5(210).
39. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Bmj 2015;351:h5527.
40. Moher D, Liberati, A., Tetzlaff, J., & Altman, D. G.,. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. . PLoS Medicine, 6(7), 2009;6(7).
41. Xu B, Al Rasheed Mohamed Rizwan H, Antonescu Cristina R, Alex D, Frosina D, Ghossein R, et al. Pan-Trk immunohistochemistry is a sensitive and specific ancillary tool for diagnosing secretory carcinoma of the salivary gland and detecting ETV6-NTRK3 fusion. Histopathology 2020;76(3):375-82.
42. Zhao M, Yin M, Kuick CH, Chen H, Aw SJ, Merchant K, et al. Congenital mesoblastic nephroma is characterised by kinase mutations including EGFR internal tandem duplications, the ETV6-NTRK3 fusion, and the rare KLHL7-BRAF fusion. Histopathology 2020;77(4):611-21.
43. Bell D, Ferrarotto R, Liang L, Goepfert RP, Li J, Ning J, et al. Pan-Trk immunohistochemistry reliably identifies ETV6-NTRK3 fusion in secretory carcinoma of the salivary gland. Virchows Arch 2020;476(2):295-305.
44. Brčić I, Godschachner Theresa M, Bergovec M, Igrec J, Till H, Lackner H, et al. Broadening the spectrum of NTRK rearranged mesenchymal tumors and usefulness of pan-TRK immunohistochemistry for identification of NTRK fusions. Mod Pathol 2021;34(2):396-407.
45. Choi Y, Won YJ, Lee S, Kim A, Kim Y, Park WY, et al. Cytoplasmic TrkA Expression as a Screen for Detecting NTRK1 Fusions in Colorectal Cancer. Transl Oncol 2018;11(3):764-70.
46. Elfving H, Brostrom E, Moens LNJ, Almlof J, Cerjan D, Lauter G, et al. Evaluation of NTRK immunohistochemistry as a screening method for NTRK gene fusion detection in non-small cell lung cancer. Lung Cancer 2021;151:53-9.
47. Fu Y, Li Z, Gao F, Yang J, Wu H, Zhang B, et al. MLH1/PMS2 Expression Could Tell Classical NTRK Fusion in Fluorescence In Situ Hybridization Positive Colorectal Carcinomas. Front Oncol 2021;11:669197.
48. Gatalica Z, Xiu J, Swensen J, Vranic S. Molecular characterization of cancers with NTRK gene fusions. Mod Pathol 2019;32(1):147-53.
49. Lee YC, Chen JY, Huang CJ, Chen HS, Yang AH, Hang JF. Detection of NTRK1/3 Rearrangements in Papillary Thyroid Carcinoma Using Immunohistochemistry, Fluorescent In Situ Hybridization, and Next-Generation Sequencing. Endocr Pathol 2020;31(4):348-58.
50. Rudzinski ER, Lockwood CM, Stohr BA, Vargas SO, Sheridan R, Black JO, et al. Pan-Trk Immunohistochemistry Identifies NTRK Rearrangements in Pediatric Mesenchymal Tumors. Am J Surg Pathol 2018;42(7):927-35.
51. Gambella A, Senetta R, Collemi G, Vallero SG, Monticelli M, Cofano F, et al. NTRK Fusions in Central Nervous System Tumors: A Rare, but Worthy Target. Int J Mol Sci 2020;21(3):23.
52. Hsiao SJ, Zehir A, Sireci AN, Aisner DL. Detection of Tumor NTRK Gene Fusions to Identify Patients Who May Benefit from Tyrosine Kinase (TRK) Inhibitor Therapy. J Mol Diagn 2019;21(4):553-71.
53. Solomon JP, Hechtman JF. Detection of NTRK Fusions: Merits and Limitations of Current Diagnostic Platforms. Cancer Res 2019;79(13):3163-8.
54. Wong D, Yip S, Sorensen PH. Methods for Identifying Patients with Tropomyosin Receptor Kinase (TRK) Fusion Cancer. Pathol Oncol Res 2019;29:29.
55. Solomon JP, Benayed R, Hechtman JF, Ladanyi M. Identifying patients with NTRK fusion cancer. Ann Oncol 2019;30 Suppl 8:viii16-viii22.
56. Hawkins RC. Laboratory turnaround time. Clin Biochem Rev 2007;28(4):179-94.

57. Ministry of Health and Care Services. A white paper on priority setting in the Norwegian health care sector. 2016. Meld. St. 34 (2015–2016). Tilgjengelig fra: https://www.regjeringen.no/no/dokumenter/meld.-st.-34-20152016/id2502758/
58. Shabaruddin FH, Fleeman ND, Payne K. Economic evaluations of personalized medicine: existing challenges and current developments. Pharmgenomics Pers Med 2015;8:115-26.
59. Hatz MH, Schremser K, Rogowski WH. Is individualized medicine more cost-effective? A systematic review. Pharmacoeconomics 2014;32(5):443-55.
60. Kasztura M, Richard A, Bempong NE, Loncar D, Flahault A. Cost-effectiveness of precision medicine: a scoping review. Int J Public Health 2019;64(9):1261-71.
61. National Institute for Health and Clinical Excellence. Diagnostic Assessment Programme Manual. 2011.
62. Annemans L, Redekop K, Payne K. Current methodological issues in the economic assessment of personalized medicine. Value Health 2013;16(6 Suppl):S20-6.
63. Fourth, red. Drummond M, Sculpher M, M C, K. , Stoddart G, Torrance G. Methods for the Economic Evaluation of Health Care Programmes: Oxford University Press; 2015.
64. Kievits T, Niese D, L.T. H, P. C, Gledic S, Roediger A, et al. Personalized Medicine:Status qua and Challenges. The European Association for Bioindustries; Bioindustries TEAf; 201.
65. Cancer Registry of Norway. Cancer in Norway. Cancer Registry of Norway; 2020.
66. Straub A, Linz C, Strobel S, Hartmann S, Hohm J, Fuchs A, et al. Mammary analogue secretory carcinoma of a salivary gland of the hard palate with contralateral cervical lymph node metastases: A case report. Mol Clin Oncol 2021;15(5):226.
67. Pohlodek K, Meciarova I, Grossmann P, Martinek P, Kinkor Z. Secretory carcinoma of the breast: A case report. Int J Surg Case Rep 2019;56:74-7.
68. Novochadlo Kluppel E, Rodrigues da Costa L, Marquetto Tognolo C, do Nascimento A, Grignet Ribeiro M, Girardi Fachin C. Secretory breast carcinoma in a male child: Case report and literature review. Int J Surg Case Rep 2020;73:310-4.
69. Burningham Z, Hashibe M, Spector L, Schiffman JD. The epidemiology of sarcoma. Clin Sarcoma Res 2012;2(1):14.
70. Kimani W, Ashiundu E, Saula P, Kimondo M, Keitany K. Congenital mesoblastic nephroma: Case study. Journal of Pediatric Surgery Case Reports 2020;55(101336).
71. Secretariat of the Special Committee on Beating Cancer. Backgrounf note on paediatric and rare cancers. European Parliament; 2021.
72. Janssen E, Stavanger University Hospital. Personal communication regarding the costs associated to diagnostic methods for the detection of ROS1. 2020-2021.
73. Pennell NA, Mutebi A, Zhou ZY, Ricculli ML, Tang WX, Wang HL, et al. Economic Impact of Next-Generation Sequencing Versus Single-Gene Testing to Detect Genomic Alterations in Metastatic Non-Small-Cell Lung Cancer Using a Decision Analytic Model. Jco Precision Oncology 2019;3.
74. Pietrantonio F, Di Nicolantonio F, Schrock AB, Lee J, Tejpar S, Sartore-Bianchi A, et al. ALK, ROS1, and NTRK rearrangements in metastatic colorectal cancer. J Natl Cancer Inst 2017;109 (12) (no pagination)(djx089).
75. Forskrift om endring i forskrift om godtgjørelse for å yte poliklinisk helsehjelp i spesialisthelsetjenesten (poliklinikkforskriften). Tilgjengelig fra: https://lovdata.no/dokument/LTI/forskrift/2020-12-22-3092
76. Norwegian Medicine Agency. Retningslinjer for dokumentasjonsgrunnlag for hurtig metodevurdering av legemidler: Gyldig fra 01.01.2018. Oppdatert 20.05.2020. Oslo: Norwegian Medicine Agency; 2017. Tilgjengelig fra: https://legemiddelverket.no/Documents/Offentlig%20finansiering%20og%20pris/Dokumentasjon%20til%20metodevurdering/Retningslinjer%2020.05.2020.pdf

77. Hetchman JF. NTRK insights: best practices for pathologists. Mod Pathol 2021.
78. Burns PB, Rohrich R, Chung CC. The Levels of Evidence and their role in Evidence-Based Medicine. Plast Reconstr Surg 2021;128(1):305–10.
79. Entrectinib for the treatment of NTRK fusion-positive solid tumours (ID1512) [ongoing project]. EUnetHTA POP Database: National Institute for Health & Care Excellence [lest]. Tilgjengelig fra: https://www.nice.org.uk/guidance/indevelopment/gid-ta10414
80. DNV GL.-BigMed. Implementing NGS-based diagnostics in cancer care: Technical and organizational factors in the Nordics. 2021.
81. San Miguel L, Hulstaert F. The importance of test accuracy in economic evaluations of companion diagnostics. J Comp Eff Res 2015;4(6):569-77.
82. Faulkner E, Annemans L, Garrison L, Helfand M, Holtorf AP, Hornberger J, et al. Challenges in the development and reimbursement of personalized medicine-payer and manufacturer perspectives and implications for health economics and outcomes research: a report of the ISPOR personalized medicine special interest group. Value Health 2012;15(8):1162-71.
83. Schluckebier L, Caetano R, Garay OU, Montenegro GT, Custodio M, Aran V, et al. Cost-effectiveness analysis comparing companion diagnostic tests for EGFR, ALK, and ROS1 versus next-generation sequencing (NGS) in advanced adenocarcinoma lung cancer patients. BMC Cancer 2020;20(1):875.
84. Meadows NA, Morrison A, Brindley DA, Schuh A, Barker RW. An evaluation of regulatory and commercial barriers to stratified medicine development and adoption. Pharmacogenomics J 2015;15(1):6-12.
85. The European parliament and of the council of the European Union. Regulation (EU) 2017/746 of the European parliament and of the council of 5 April 2017 on in vitro diagnostic medical devices. 2017.
86. Anderson J, Gibson S, Sebire NJ. Expression of ETV6-NTRK in classical, cellular and mixed subtypes of congenital mesoblastic nephroma. Histopathology 2006;48(6):748-53.
87. Brisudova A, Skarda J. Gene rearrangement detection by next-generation sequencing in patients with non-small cell lung carcinoma. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 2020;164(2):127-32.
88. Hechtman JF, Benayed R, Hyman DM, Drilon A, Zehir A, Frosina D, et al. Pan-Trk Immunohistochemistry Is an Efficient and Reliable Screen for the Detection of NTRK Fusions. Am J Surg Pathol 2017;41(11):1547-51.
89. Pfarr N, Kirchner M, Lehmann U, Leichsenring J, Merkelbach-Bruse S, Glade J, et al. Testing NTRK testing: Wet-lab and in silico comparison of RNA-based targeted sequencing assays. Genes Chromosomes Cancer 2020;59(3):178-88.
90. Zarabi SK, Azzato EM, Tu ZJ, Ni Y, Billings SD, Arbesman J, et al. Targeted next generation sequencing (NGS) to classify melanocytic neoplasms. J Cutan Pathol 2020;47(8):691-704.
91. Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, et al. Anchored multiplex PCR for targeted next-generation sequencing. Nat Med 2014;20(12):1479-84.

Appendices

Appendix 1 NTRK gene fusions identified in adult and paediatric cancers by relative frequency of NTRK gene fusions

	Fusion partner		
	NTRK1	NTRK2	NTRK3
Adult cancers			
High frequency (>80%)			
Mammary analogue secretory carcinomas			ETV6
Secretory breast carcinoma			ETV6
Intermediate frequency (5%–25%)			
Papillary thyroid cancer	TFG, SSBP2, SQSTM1, TPR, PPL		ETV6, RBPMS
Low frequency (<5%)			
Appendiceal cancer	LMNA		
Glioma/glioblastoma	ARHGEF2, BCAN, CHTOP, NFASC,	BCR, AFAP1, SQSTM1	AFAP1, ZNF710, EML4
Astrocytoma		QK1, NACC2	
Gastrointestinal stromal tumour			ETV6
Head and neck cancer		PAN3	LYN
Lung cancer	CD74, GRPAP1, IRF2BP2, MPRIP, P2RY8, SQSTM1, TPM3	TRIM24	
Sarcoma	TPM3, LMNA		TPM4
Breast cancer	CGN, GATAD2B, LMNA, MDM4, PEAR1, TPM3,		ETV6

Acute lymphoblastic leukaemia, acute myeloid leukaemia, histiocytosis, multiple myeloma, dendritic cell neoplasms			ETV6
Uterine sarcoma	LMNA, TPM3, TPR		RBPMS
Cholangiocarcinoma	LMNA, RABGAP1L		
Pancreatic cancer	CTRC10		
Melanoma	DDR2, GON4L, TRIM63	TRAF2	ETV6
Colorectal cancer	LMNA, TPM3, SCYL3		ETV6
Paediatric cancers			
High frequency (>80%)			
Secretory breast carcinoma			ETV6
Infantile fibrosarcoma and other mesenchymal tumours	SQSTM1, TPM3, LMNA		EML4, ETV6
Cellular and mixed congenital mesoblastic nephroma	TPR, LMNA		EML4, ETV6
Intermediate frequency (5%–25%)			
Papillary thyroid cancer	TPR, IRF2BP2, TPM3		ETV6
Spitz tumours	TP53, LMNA		ETV6, MYH9, MYO5A
Paediatric high-grade gliomas	TPM3	AGBL4, VCL	ETV6, BTB1
Low frequency (<5%)			
Ganglioglioma		TLE	
Astrocytoma		NACC2, QK1	

Source: (19)

Appendix 2 Glossary

Analytical validity	or technical performance, is a test's ability to accurately and reliably *measure* a biomarker of interest (sensitivity, specificity, assay robustness, and quality control). According to EGAPP definitions https://www.cdc.gov/genomics/gtesting/egapp/recommend/method.htm
Antibody clone	selected antibodies with the characteristics that work well for immunohistochemistry (IHC)
Chemotherapy	a drug treatment aimed at killing cancer-cells
Clinical utility	impact on patient outcomes, refers to how likely it is that using the test to guide clinical decisions will significantly improve outcomes related to patients health and well-being (benefits vs. harms, whether using the tests gives added value to not using it, effectiveness, and efficacy). According to EGAPP definitions https://www.cdc.gov/genomics/gtesting/egapp/recommend/method.htm.
Clinical validity	or the strength of clinical correlation, is a test's ability to accurately and reliably identify or *predict* the disorder of interest (sensitivity, specificity, positive predictive value, negative predictive value). According to EGAPP definitions https://www.cdc.gov/genomics/gtesting/egapp/recommend/method.htm
Colorectal carcinomas	a cancer, or malignant tumour, of the large intestine, which may affect the colon or rectum.
Concordance rate	the number of test results that are concordant (in agreement) over the total number of tests analysed
Confidence interval	a type of estimate computed from the statistics of the observed data that proposes a range of plausible values for an unknown parameter (e.g., the mean)
Congenital mesoblastic nephroma	a rare tumour with an estimated incidence of about 8/million in children under 15 years of life. It is, however, the most common renal tumour in neonates, with more than 80% presenting in the neonatal period.
DNA-based NGS	NGS method that used deoxyribonucleic acid in the analysis
End-to-end study	a study that follows patients from testing, through treatment, to final outcomes
Entrectinib	is a medication for the treatment of cancer. It is a selective tyrosine kinase inhibitor of TRK A, B and C, C-ros oncogene 1 (ROS1) and anaplastic lymphoma kinase (ALK)
FoundationOne CDx assay (F1CDx)	a NGS based in vitro diagnostic device that is capable of detecting several mutations in addition to NTRK gene fusions

Fluorescence in situ hybridisation (FISH)	a laboratory method for detecting and locating a specific DNA sequence on a chromosome that relies on exposing chromosomes to a small DNA sequence called a probe that has a fluorescent molecule attached to it; the probe sequence binds to its corresponding sequence on the chromosome.
Fusion gene	a gene made by joining parts of two different genes
Gene alteration	a somatic gene alteration
Genomics	a term that refers to the molecular composition of a tumour
Immuno-Histo-Chemistry (IHC)	a laboratory method that uses antibodies to check for certain antigens in tissue samples. The antibodies are usually linked to an enzyme or a fluorescent dye, which when activated allows the antigen to be seen under a microscope
Immunotherapy	a type of cancer treatment that helps your immune system fight cancer
Indeterminate	results that are neither positive nor negative
Infantil fibrosarkoma	a type of cancer that forms in fibrous (connective) tissue. Infantile fibrosarcoma usually occurs in infants and young children
Larotrectinib	is a medication for the treatment of cancer. It is an inhibitor of tropomyosin kinase receptors TrkA, TrkB, and TrkC.
Molecular test	a laboratory test that checks for certain genes, proteins, or other molecules in a sample of tissue, blood, or other body fluid, or that check for certain changes in a gene or chromosome
Multi-gene panel	a genetic test that uses next-generation sequencing to test multiple genes simultaneously
Multiplexing	A method for detecting multiple genetic alteration simultaneously
Mutation (somatic)	a genetic alteration acquired by a cell that can be passed to the progeny of the mutated cell in the course of cell division
Narrative review	a narrative (non-systematic) which purpose is to identify a few studies that describe a problem of interest
Next Generation Sequencing (NGS)	also called massively parallel, deep sequencing or multigene panel, is a DNA sequencing technology by which entire human genome can be sequenced within a single day
Non-small-cell lung carcinoma (NSCLC)	any type of epithelial lung cancer other than small-cell lung carcinoma (SCLC), which accounts for about 85% of all lung cancers
Osteosarcoma	Osteosarcoma (also called osteogenic sarcoma) is the most common type of cancer that starts in the bones.

Pan Trk	Pan-TRK (clone EPR17341) is directed against the C-terminal region of TRK (tropomyosin receptor kinase) A, B, and C proteins, which are encoded by NTRK1, NTRK2, and NTRK3 genes respectively. Pan-TRK IHC staining is a useful screen for identification of NTRK protein overexpression caused by gene fusions.
Papillary Thyroid Carcinoma	Papillary carcinoma (PTC) is the most common form of well-differentiated thyroid cancer, and the most common form of thyroid cancer to result from exposure to radiation
Mesenchymal tumours	Mesenchymal tissue neoplasms are soft tissue tumours, also known as connective tissue tumours
Radiation therapy	a type of cancer treatment that uses high energy beams most often X-rays, but also protons or other types of energy, to kill cancer cells
RNA-based NGS	NGS-method that used ribonucleic acid for the analysis
Gene rearrangement	a programmed DNA recombination event that occurs during cellular differentiation to reconstitute a functional gene from gene segments separated in the genome
Reverse transcription polymerase chain reaction	or RT-PCR, is a laboratory technique combining reverse transcription of RNA into DNA and amplification of specific DNA targets using polymerase chain reaction
Secretory carcinoma	Secretory carcinoma is a recently described malignancy affecting the salivary glands of the head and neck
Sensitivity	the ability of a test to correctly identify those with the disease (true positive rate)
Single-gene test	a test that looks for changes in one gene at the time
Solid tumour	an abnormal mass of tissue that usually does not contain cysts or liquid areas.
Specificity	the ability of the test to correctly identify those without the disease (true negative rate)
Somatic gene mutation	a type of alteration in DNA that occurs after conception.
Systematic review	a review of the evidence on a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyse data from the studies that are included in the review
Turnaround time (TAT)	the total test cycle which includes ordering of test, collection, identification, transportation, preparation, analysis, reporting, interpretation and action

Therapeutic TAT	the interval between when a test is requested to the time a treatment decision is made
TrkA	a receptor in the tyrosine protein kinase family

Appendix 3 Full Search strategy

Database	Search hits
Cochrane Central Register of Controlled Trials (Wiley) - April 2020	8
Embase (Ovid) – April 2020 + May 2021	1713
MEDLINE (Ovid) – April 2020 + May 2021	1275
ClinicalTrials.gov (US National Institutes of Health) – April 2020 + May 2021	64
International Clinical Trials Registry Platform (WHO) – May 2021 (not searched April 2020 due to database technical issues)	22
PROSPERO International Prospective Register of Systematic Reviews (National Institute for Health Research, UK) - April 2020	1
EUnetHTA POP database (EUnetHTA) - April 2020	3
Number of references imported to EndNote	3086
Number of references exported to screening (after deduplication)	2379

Search strategies

Database: Cochrane Central Register of Controlled Trials Search date: 2020-04-02 Search interface: Advanced search – Search manager	
#1	(([mh ^"Receptor Protein-Tyrosine Kinases"] OR [mh ^"Receptor, trkA"] OR [mh ^"Receptor, trkB"] OR [mh ^"Receptor, trkC"]) AND ([mh ^"Oncogene Proteins, Fusion"] OR [mh ^"Gene Fusion"])) OR ((neurotroph* OR NTRK1 OR MTC OR TRK OR TRK1 OR TRKA OR "Trk-A" OR "p140-TrkA" OR "p140(trkA)" OR NTRK2 OR EIEE58 OR "GP145-TrkB" OR "gp145(trkB)" OR OBHD OR TRKB OR "trk-B" OR NTRK3 OR "GP145-TrkC" OR TRKC OR "gp145(trkC)" OR "trk-C" OR NTRK OR "NTRK1/2/3" OR "NTRK1,2,3" OR "NTRK1-3" OR "TRKA/B/C") NEAR/6 fusion*)

Database: Embase <1974 to 2020 April 01> + <1974 to 2021 May 17> Ovid MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations, Daily and Versions(R) 1946 to April 01, 2021 + Ovid MEDLINE(R) ALL <1946 to May 17, 2021> Search date: 2020-04-02 + 2021-05-18 Search interface: Advanced search	
1	((Receptor Protein-Tyrosine Kinases/ or Receptor, trkA/ or Receptor, trkB/ or Receptor, trkC/) and (Oncogene Proteins, Fusion/ or Gene Fusion/)) use medall or ((Protein Tyrosine Kinase/ or Protein Tyrosine Kinase A/ or Brain Derived Neurotrophic Factor Receptor/ or Neurotrophin 3 Receptor/) and Gene Fusion/) use oemezd or ((neurotroph* or NTRK1 or MTC or TRK or TRK1 or TRKA or Trk-A or p140-TrkA or "p140(trkA)" or NTRK2 or EIEE58 or GP145-TrkB or "gp145(trkB)" or OBHD or TRKB or trk-B or NTRK3 or GP145-TrkC or TRKC or "gp145(trkC)" or trk-C or NTRK or "NTRK1/2/3" or "NTRK1,2,3" or "NTRK1-3" or "TRKA/B/C") adj6 fusion*).tw,kw,kf.
2	(exp Animals/ not Humans/) use medall or (exp Animal/ not Human/) use oemezd or (animal* or mouse or mice or murine or rat or rats).ti.
3	(Conference Abstract or Conference Review or Conference Paper).pt use oemezd
4 (April 2020)	1 not 2 [ppezv + oemezd]

4 (May 2021)	1 not (2 or 3) [medall + oemezd]

Database: NIH Clinical Trials (clinicaltrials.gov) Search date: 2020-04-02 + 2021-05-18 Search interface: standard søk (Find a study > Other terms)
((NTRK OR NTRK1 OR NTRK2 OR NTRK3 OR TRK OR TRKA OR TRKB OR TRKC) AND fusion)

Database: International Clinical Trials Registry Platform (WHO) Search date: 2021-05-18 Search interface: standard search
NTRK AND fusion OR NTRK1 AND fusion OR NTRK2 AND fusion OR NTRK3 AND fusion OR TRK AND fusion OR TRKA AND fusion OR TRKB AND fusion OR TRKC AND fusion

Database: PROSPERO International Prospective Register of Systematic Reviews (National Institute for Health Research) Search date: 2020-04-02	
1	((neurotroph* OR NTRK1 OR MTC OR TRK OR TRK1 OR TRKA OR "Trk-A" OR "p140-TrkA" OR "p140(trkA)" OR NTRK2 OR EIEE58 OR "GP145-TrkB" OR "gp145(trkB)" OR OBHD OR TRKB OR "trk-B" OR NTRK3 OR "GP145-TrkC" OR TRKC OR "gp145(trkC)" OR "trk-C" OR NTRK OR "NTRK1/2/3" OR "NTRK1,2,3" OR "NTRK1-3" OR "TRKA/B/C") AND fusion*)

Database: EUnetHTA POP database Search date: 2020-04-02 Search interface: Search
NTRK fusion* [Search for keywords: all (AND)]

Appendix 4 List of excluded studies

Study First author (reference no.)	Cause for exclusion of study
Anderson 2006 (86)	Not a comparison between different types of analytical tests
Brisudova 2020 (87)	Focus not on NTRK fusions
Hechtman 2017 (88)	Unfinished/intermediate study. Final study by Salomon 2019.
Pfarr 2020 (89)	Not a comparison between different types of analytical tests
Xu 2020 (41)	Unclear comparison/results. No clarification received from authors (1 reminder e-mail sent)
Zarabi 2020 (90)	Results not specific for NTRK fusions/NTRK not in focus
Zhao 2020 (42)	Unclear comparison/results. No clarifications received from authors (1 reminder e-mail sent)
Zheng 2014 (91)	Results not specific for NTRK fusions/NTRK not in focus

Appendix 5 Description of the EGAPP quality assessment tool

Table 5.1. Hierarchies of data sources and study designs for the components of evaluation

Level [a]	Analytic validity	Clinical validity	Clinical utility
1	Collaborative study using a large panel of well characterized samples Summary data from well-designed external proficiency testing schemes or interlaboratory comparison programs	Well-designed longitudinal cohort studies Validated clinical decision rule[b]	Meta-analysis of randomized controlled trials (RCT)
2	Other data from proficiency testing schemes Well-designed peer-reviewed studies (e.g., method comparisons, validation studies) Expert panel reviewed FDA summaries	Well-designed case-control studies	A single randomized controlled trial
3	Less well-designed peer-reviewed studies	Lower quality case-control and cross-sectional studies Unvalidated clinical decision rule[b]	Controlled trial without randomization Cohort or case-control study
4	Unpublished and/or non-peer reviewed research, clinical laboratory, or manufacturer data Studies on performance of the same basic methodology, but used to test for a different target	Case series Unpublished and/or non-peer reviewed research, clinical laboratory or manufacturer data Consensus guidelines Expert opinion	Case series Unpublished and/or non-peer reviewed studies Clinical laboratory or manufacturer data Consensus guidelines Expert opinion

[a] Highest level is 1

[b] A clinical decision rule is an algorithm leading to result categorization. It can also be defined as a clinical tool that quantifies the contributions made by different variables (e.g., test result, family history) in order to determine classification/interpretation of a test result (e.g., for diagnosis, prognosis, therapeutic response) in situations requiring complex decision-making.[1]

[1] Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: testing for cytochrome P450 (*CYP450*) polymorphisms in adults with nonpsychotic depression treated with selective serotonin reuptake inhibitors. *Genet Med* 2007; 9:819–825.

Table 5.2. Criteria for assessing quality of individual studies (internal validity)[1]

Analytic validity	Clinical validity	Clinical utility
Adequate descriptions of the index test (test under evaluation) Source and inclusion of positive and negative control materials Reproducibility of test results Quality control/assurance measures *Adequate descriptions of the test under evaluation* Specific methods/platforms evaluated Number of positive samples and negative controls tested *Adequate descriptions of the basis for the "right answer"* Comparison to a "gold standard" referent test Consensus (e.g., external proficiency testing) Characterized control materials (e.g., NIST, sequenced) *Avoidance of biases* Blinded testing and interpretation Specimens represent routinely analyzed clinical specimens in all aspects (e.g., collection, transport, processing) Reporting of test failures and uninterpretable or indeterminate results *Analysis of data* Point estimates of analytic sensitivity and specificity with 95% confidence intervals Sample size/power calculations addressed	*Clear description of the disorder/phenotype and outcomes of interest* Status verified for all cases Appropriate verification of controls Verification does not rely on *index test* result Prevalence estimates are provided *Adequate description of study design and test/methodology* *Adequate description of the study population* Inclusion/exclusion criteria Sample size, demographics Study population defined and representative of the clinical population to be tested Allele/genotype frequencies or analyte distributions known in general and subpopulations *Independent blind comparison with appropriate, credible reference standard(s)* Independent of the test Used regardless of test results Description of handling of indeterminate results and outliers Blinded testing and interpretation of results *Analysis of data* Possible biases are identified, and potential impact discussed Point estimates of clinical sensitivity and specificity with 95% confidence intervals Estimates of positive and negative predictive values	*Clear description of the outcomes of interest* What was the relative importance of outcomes measured; which were prespecified primary outcomes and which were secondary? *Clear presentation of the study design* Was there clear definition of the specific outcomes or decision options to be studied (clinical and other endpoints)? Was interpretation of outcomes/endpoints blinded? Were negative results verified? *Was data collection prospective or retrospective?* If an experimental study design was used, were subjects randomized? Were intervention and evaluation of outcomes blinded? Did the study include comparison with current practice/empirical treatment (value added)? *Intervention* What interventions were used? What were the criteria for the use of the interventions? *Analysis of data* Is the information provided sufficient to rate the quality of the studies? Are the data relevant to each outcome identified? Is the analysis or modeling explicit and understandable? Are analytic methods prespecified, adequately described, and appropriate for the study design? Were losses to follow-up and resulting potential for bias accounted for? Is there assessment of other sources of bias and confounding? Are there point estimates of impact with 95% CI? Is the analysis adequate for the proposed use?

[1] Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: testing for cytochrome P450 (*CYP450*) polymorphisms in adults with nonpsychotic depression treated with selective serotonin reuptake inhibitors. *Genet Med* 2007; 9:819–825.

Table 5.3. Grading the quality of evidence for the individual components of the chain of evidence (key questions)[2]

Adequacy of information to answer key questions	Analytic validity	Clinical validity	Clinical utility
Convincing	*Studies that provide confident estimates of analytic sensitivity and specificity using intended sample types from representative populations* Two or more Level 1 or 2 studies that are generalizable, have a sufficient number and distribution of challenges, and report consistent results One Level 1 or 2 study that is generalizable and has an appropriate number and distribution of challenges	*Well-designed and conducted studies in representative population(s) that measure the strength of association between a genotype or biomarker and a specific and well-defined disease or phenotype* Systematic review/meta-analysis of Level 1 studies with homogeneity Validated Clinical Decision Rule High Quality Level 1 cohort study	*Well-designed and conducted studies in representative population(s) that assess specified health outcomes* Systematic review/meta-analysis of randomized controlled trials showing consistency in results At least one large randomized controlled trial (Level 2)
Adequate	Two or more Level 1 or 2 studies that Lack the appropriate number and/or distribution of challenges Are consistent, but not generalizable Modeling showing that lower quality (Level 3, 4) studies may be acceptable for a specific well- defined clinical scenario	Systematic review of lower quality studies Review of Level 1 or 2 studies with heterogeneity Case/control study with good reference standards Unvalidated Clinical Decision Rule (Level 2)	Systematic review with heterogeneity One or more controlled trials without randomization (Level 3) Systematic review of Level 3 cohort studies with consistent results
Inadequate	Combinations of higher quality studies that show important unexplained inconsistencies One or more lower quality studies (Level 3 or 4) Expert opinion	Single case-control study Nonconsecutive cases Lacks consistently applied reference standards Single Level 2 or 3 cohort/case-control study Reference standard defined by the test or not used systematically Studies not blinded Level 4 data	Systematic review of Level 3 quality studies or studies or studies with heterogeneity Single Level 3 cohort or case-control study Level 4 data

[2] Sawaya GF, Guirguis-Blake J, LeFevre M, et al. Update on methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. *Ann Intern Med* 2007; 147:871– 875.

Appendix 6 Technical details of evaluated tests

Author / Year	IHC	FISH		PCR				RNA-NGS		DNA-NGS	
	Antibody clone	Dual Break apart	Fusion	RT-PCR[1]	Defined gene partners	3' /5' NTRK ratio	RT- PCR[2]	Hybridization capture	Amplicon (anchored multiplex RT-PCR)	Hybridization capture	Amplicon (target enrichment by PCR)
Bell 2020	X	X		X							
Brcic 2020	X								X		
Choi 2018	X	X									
Elfving 2021	X							X			
Fu 2021	X	X							X		X
Gatalica 2019	X									X	
Lee 2020	X	X						X		X	
Rudzinski 2018	X									X	
Salomon 2020	X							X		X	

[1] RT- PCR: Reverse Transcriptome Polymerase Chain Reaction; DNA: Deoxyribonucleic acid; IHC: immunohistochemistry; MA: Massachusetts; MD (Rockville): Maryland; MSK: Memorial Sloan Kettering; NGS: next-generation sequencing; RT-PCR: real time polymerase chain reaction; RNA: ribonucleic acid; 2 RT-PCR: real time – polymerase chain reaction; UCSF: University of California San Francisco; UK: United Kingdom; UW: University of Washington

	Bell 2020	Brcic 2020	Choi 2018	Elfving 2021	Fu 2021	Gatalica 2019	Lee 2020	Rudzinski 2018	Salomon 2020
Type of samples	FFPE (also FF for PCR)	FFPE/TMAs	FFPE	FFPE/TMAs	FFPE	FFPE	FFPE/TMAs	NR	FFPE
Amount of material used (e.g., in ng or no of slices)									
IHC	NR	4 µm TMA sections (with 4 0.6 mm cores)	NR	4 µm TMA sections (with two 1 mm cores)	4-µm TMA (10%) sections	NR	4-µm TMA sections (one 1 mm core)	NR	NR
FISH	NR	NA	NR	NA	NR	NA	NR	NA	NA
RT-PCR	3 10-µm unstained FFPE sections	NA	NA	NA	NA	NA	NR	NA	NA
NGS	NA	250 ng total RNA (5–8 × 10 µm FFPE sections)	NA	100 ng RNA (due to heavily degraded samples)	NR	NR	NR	NR	NR
Library preparation method (NGS only)									
	NA	Ion Torrent Proton using the Ion PI IIi-Q Sequencing 200kit (Thermo Fischer, MA)	NA	According to the TruSight Tumor 170 reference guide (Illumina, San Diego, CA, USA).	Archer Universal RNA Reagent Kit v2 (ArcherDx, Boulder, CO). Library sequencing was done using a MiSeqDx instrument (Illumina, San Diego, CA). DNA-based performed by Macrogen USA (Rockville,MD).	NGS constructed using ArcherDx Fusion Plex Assay (ArcherDX, CO); RUO	ArcherDX Fusion-Plex Comprehensive Thyroid and Lung Panel (ArcherDX, Boulder, CO, USA): RUO	NR	NR
Platforms/systems used for analysis (NGS only)									

	Bell 2020	Brcic 2020	Choi 2018	Elfving 2021	Fu 2021	Gatalica 2019	Lee 2020	Rudzinski 2018	Salomon 2020
	NA	Ion Torrent Proton using the Ion PI Hi-Q Sequencing 200 kit (Thermo Fischer, Waltham, MA).	NA	Fusion analysis was performed by TruSight Tumor 170 v2.0 Local App (Illumina, San Diego, CA, USA) and Arriba v1.1.0 (https://github.com/suhrig/arriba/).	Archer Analysis Pipeline Virtual Machine (https://archerdx.com).DNA Ion Torrent (Life Technologies/Thermo FisherScientific, Waltham, MA) platform.	Platforms/systems used for analysis (NGS only)	Archer Analysis bioinformatics platform.	UW Oncoplex, a targeted DNA-based platform; UCSF500 Cancer Gene Panel, a targeted DNA-based platform; platforms chosen locally (and 1 was confirmed by ETV6 FISH)	RNA Sequenced via anchored multiple PCR (via the Archer platform)
Software used									
	NA	AnchorDX Analysis (software) 5.1.3	NA	NA	DNA Bioinformatics analysis of NGS data was processed by Torrent Server Suite 4.2	NR	ArcherDX FusionPlex; RUO.	UW Oncoplex (NF); UCSF 500 gene panel test	NR
Test and Regulatory Status (EMA, CE, RUO and/or FDA approved)									
IHC	Pan-Trk rabbit monoclonal antibody [EPR17341] (Abcam, MA): RUO	Pan Trk rabbit monoclonal antibody [EPR17341] Roche, Ventana): CE-IV marked	Anti TrkA C-Terminal monoclonal antibody (TA806413, ORiGene, US): EU commission.	Ventana pan TRK antibody [EPR17341] (Roche, Basel, Switzerland); RUO	Pan-Trk antibody [EPR17341] (Abcam, USA); RUO	Pan-Trk rabbit monoclonal antibody [EPR17341] (Abcam, MA): RUO;	Anti-pan-TRK antibody (clone EPR17341, Abcam, Cambridge, MA, USA); RUO	Anti-TrkA monoclonal antibody clone EP1058Y (Abcam, Cambridge, MA) and a pan-Trk monoclonal antibody clone EPR17341 (Abcam) RUO	Pan-Trk antibody clone EPR17341 (Abcam, Cambridge, MA); RUO
FISH	Cytocell FISH ETV6-NTRK3 dual-color probes (Oxford Gene Technology (UK)); RUO	NA	Commercially available split FISH probes used for detecting NTRK1	NA	NTRK1/2/3 Dual Color Break Apart Probe (Anbiping, China) NF	NA	Dual-color break-apart FISH probes for NTRK1 and NTRK3 (ZytoVision, Bremerhaven, Germany), and ETV6	NA	NA

	Bell 2020	Brcic 2020	Choi 2018	Elfving 2021	Fu 2021	Gatalica 2019	Lee 2020	Rudzinski 2018	Salomon 2020
							(ZytoVision) ZytoLight ® SPEC NTRK1|3 Dual Color Break Apart Probe CE and IVD approved in certain countries		
RT-PCR	Super Script III First-Strand One-Step RT-PCR system (Life Technologies).PureLink FFPE Total RNA Isolation Kit (Life Technologies, Carlsbad, CA): RUO and RNeasy Plus Universal Mini Kit (Qiagen, Hilden, Germany); NA	NA	NA	NA	NA	NA	NA	NA	NA
NGS	NA	Ion Pi Ili-Q sequencint 200 kit (Thermo Fischer, MA): RUO	NA	Illumina NextSeq High Output v2 kits on a NextSeq 550 system. RUO	DNA or RNA FFPE Kit (Promega, Madison,WI); MiSeqDx System is FDA-regulated, CE-IVD-marked for IVD testing. Archer Universal DNA reagent Kit 2: RUO	SureSelext XT, RUO NextSeg; RUO	Sample sequenced on Illumina NextSeq 500 sequencer (Illumina, San Diego, US); RUO	UW Oncoplex, NF: UCSF500 Cancer Gene Panel (unclear) ETV6 FISH (unclear)	Targeted DNA-based NGS panel (MSK-IMPACT) FDA approved and an RNA-based NGS panel (MSK-Fusion) (unclear)
Commercial or 'in-house' test (all)									
	Commercial test(s)	Commercial	Commercial -	Commercial	Commercial	Commercial	Commercial	UW Oncoplex and UCSF 500 commercial	Pan Trk clone and MSK-IMPACT is commercial; MSK-FUSION (unclear)
Analysis									

	Bell 2020	Brcic 2020	Choi 2018	Elfving 2021	Fu 2021	Gatalica 2019	Lee 2020	Rudzinski 2018	Salomon 2020
IHC	NR	Benchmark Ultra platform with iVIEW DAB Detection Kit (Ventana Medical Systems, Tucson, AZ)	DAPI counterstaining (32–804,831; Abbott, Chicago, IL, US)	OptiView DAB kit	EnVisionTMDetection Kit, Dako, Glostrup, Denmark	Benchmark, Ventana Medical Systems,Inc. and DAKO Autostainer, Agilent.	Leica Bond-Max autostainer (Leica Biosystems, Buffalo Grove, IL, USA)	Benchmark ULTRA; (Ventana Medical Systems, Tucson, AZ)	NR
FISH	ETV6-NTRK3 Dual Fusion/Translocation FISH Probe Kit: CE approval	NA	NTRK1 Split FISH probe (FS0024; Abnova, Taiwan): RUO	NA	NR	NA	NR	NA	NA
RT-PCR	Super Script III First-Strand One-Step RT-PCR system: RUO; HotStar PCR Master Mix (Qiagen); NA	NA	NA	NA	NA	NA	NA	NA	NA
NGS	NA	Archer FusionPlex Sarcoma Panel	NA	RNA Screen Tape on a 2200 Tape Station system (Agilent, Santa Clara, CA, USA) TruSight Tumor 170 v2.0 Local App (Illumina, San Diego, CA, USA) and Arriba v1.1.0 (https://github.com/suhrig/arriba/).	Archer Analysis Pipeline VirtualMachine (https://archerdx.com). DNA Bioinformatics analysis of NGS data was processed by Torrent Server Suite 4.2	massively parallel NGS platform (592-Gene Panel utilizes SureSelect XT biotinylated RNA probes from Agilent, Santa Clara, CA) NextSeq instrument from Illumina, San Diego, CA).	Archer Analysis bioinformatics platform: NF	NR	NR
No of pathologist / analysts needed for analysis and interpretation									

	Bell 2020	Brcic 2020	Choi 2018	Elfving 2021	Fu 2021	Gatalica 2019	Lee 2020	Rudzinski 2018	Salomon 2020
	NR, but CLIA laboratories.	IHC: three pathologists reviewed all cases. All HE stained slides and IHC were re-reviewed by two soft tissue pathologists. NGS: NR	IHC: interpretation by three pathologists FISH: NR	IHC: interpretation by two blinded independent observers; NGS:NR	IHC: interpreted independently by two pathologists who were blinded to all clinical and pathological data. FISH and NGS:NR	NR.	NR	NR, but CLIA laboratories	NR, but CLIA laboratories

AZ: Arizona; CA: California; CE IV: CE European marking conformity; CLIA: Clinical laboratory improvement amendments; CO: Colorado; DNA: Deoxyribonucleic acid; FDA: Food and Drug Administration; FF: Fresh frozen; FFPE: Formalin-Fixed Paraffin-Embedded; IHC: immunohistochemical; MA: Massachusetts; MD (Rockville): Maryland; MSK: Memorial Sloan Kettering; NA: not available; NF: Not found; NGS: next-generation sequencing; NR: not reported; PCR: polymerase chain reaction; RNA: ribonucleic acid; RT-PCR: reverse transcription – polymerase chain reaction; RUO: Research use only; TMA: tissue microarray; UK: United Kingdom; US/USA: United States; UW: University of Washington

Appendix 7 Accuracy related results by cancer type (when > one study provided data)

Colorectal carcinomas

Four studies provided accuracy-related test results for colorectal carcinomas (11;45;47;48). Choi et al (45) reported high accuracy (AUROC:0.926) for IHC (as compared to FISH) for the detection of NTRK1 fusions in CRC. The number of samples positive for IHC was unclear. Fu et al (47) reported 60% consistency between IHC and RNA-based NGS, when CRC samples positive with IHC (10 of 819) was compared with NTRK positive samples by RNA-NGS (6 of 10). Gatalica et al (48) reported 50% consistency between samples positive for NTRK1 fusions with RNA-based NGS (2 of 1272), and those positive with IHC (1 of 2). Salomon and colleagues (11) reported a sensitivity of 87.5% (7/8) and a specificity of 100% (25/25) of IHC as compared to RNA based NGS. Prevalence of NTRK gene fusions in the three studies ranged from 0.2% to m 21.2%.

Lung carcinomas

Three studies provided results for NSCLC/lung cancer (11;46;48). Elfving et al (46) reported 0% consistency between IHC and RNA-based NGS when NSCLC samples NTRK positive with IHC (11 of 617) was compared with positive samples by RNA-NGS (0 of 11). Gatalica et al (48) reported 100% consistency between lung adenocarcinoma samples positive for NTRK1 and NTRK2 fusions with RNA-based NGS (1 of 4073), and those positive with IHC (1 of 1). Salomon and colleagues (11) reported a sensitivity of 87.5% (7/8) and a specificity of 100% (24/24) of IHC for detection of any NTRK fusions as compared to RNA-based NGS. Prevalence of NTRK gene fusions in the three studies ranged from 0% to 25 %.

Thyroid carcinomas

Three studies included results for PTC/Thyroid carcinomas (11;48;49). Gatalica and colleagues (48) reported 50% consistency between samples positive for NTRK3 fusions with RNA-based NGS (4 of 70 mixed thyroid carcinomas), and those positive with IHC (2 of 4 of which one was a papillary thyroid carcinoma). Lee et al. (49) reported a sensitivity of 41.7% and a specificity of 100% for IHC as compared to FISH for detecting any NTRK gene fusion in PTC. Salomon et al (11) reported a sensitivity of 81.8% (9/11) and a specificity of 100% (27/27) of IHC for detection of any NTRK fusions, as compared to RNA-based NGS. Prevalence of NTRK gene fusions in the three studies ranged from 6% to 28.9 %.

Salivary gland carcinomas

Two studies included results for salivary gland carcinomas (11;43). Bell et al (43) reported fair concordance (Cohen's kappa:0.359) between IHC and FISH for the detection of ETV6-NTRK3 fusions in three different types of salivary gland carcinomas. Salomon and colleagues (11) reported a sensitivity of 88.9% (8/9) and a specificity of 52% (13/25) of IHC for detection of any NTRK fusions, as compared to RNA-based

NGS. Prevalence of NTRK gene fusions in this study was 26.5 %. It was unclear what types of salivary gland carcinomas that were included.

Soft tissue sarcomas

Two studies reported results for soft tissue sarcomas (44;48). Bricic and colleagues (44) reported 30.2% consistency between samples positive with IHC (16 of 494) and those positive with RNA-based NGS (5 of 16). Prevalence of NTRK1 and NTRK3 in this study was 1%.

Gatalica et al (48) reported 100 % consistency between NTRK gene fusion positive samples with RNA-based NGS (1 of 478), and those positive with IHC (1 of 1). Prevalence of NTRK gene fusions was 0.2% in this study.

Appendix 8 Quality of evidence- results of the EGAPP tool

	Bell 2020 (43)	Bricic 2021 (44)	Choi 2018 (45)	Elfving 2021 (46)	Fu 2021 (47)	Galatica 2019 (48)	Lee 2020 (49)	Rudzinski 2018 (50)	Salomon 2020 (11)
Level of study Hierarchy – (grade 1-4)	3	3	3	3	3	3?	3	3	2
ANALYTICAL VALIDITY									
Adequate description of index test									
Source and inclusion of positive and negative control materials	NR	Positive control for IHC. Source given[1]	Positive and negative controls for IHC. Source given[2]	Positive control for IHC. Source given[3]	NR	Positive control for IHC. Source given[4]	Positive control for IHC. Source[5]	Positive and negative controls for IHC[6]	Positive and negative clinical cases for IHC (multiple sources)[7]
Reproducibility of test results	NR	NR	NR	NR	NR	NR	NR	NR	NR
Quality control/assurance program	NR	NR	NR	NR	NR	NR	NR	NR	NR (but CLIA)
Adequate description of the test under evaluation									
Specific methods/platforms evaluated	yes	yes	yes	yes	yes	yes	yes	yes	yes
Number of positive samples and negative controls tested	NR	NR	NR	NR	NR	NR	NR	NR (for internal controls)	Yes, for clinical samples
Adequate description of the terms for the right answer									
Comparison to a "gold standard" referent test	Yes	Partly, only for positive cases	Yes	Partly, only for positive cases	Partly, only for positive cases	Yes	Yes	Partly, only for a subset	Yes[8]
Consensus (e.g., external proficiency testing)	NR	NR	NR	NR	NR	NR	NR	NR	NR
Characterized control materials (e.g., NIST, sequenced)	NR	NR	NR	NR	NR	NR	NR	NR	**Yes**, sequenced
Avoidance of biases									
Blinded testing and interpretation	NR	NR	NR	**Yes**, for IHC interpretation	**Yes**, for IHC interpretation	NR	NR	NR	NR
Specimens represent routinely analyzed clinical specimens in all aspects (e.g., collection, transport, processing)	NR	NR	NR	no (all samples sent for deep sequencing)	NR	NR	NR	NR	NR
Reporting of test failures and uninterpretable or indeterminate results	no	no	no	**yes**	no	no	no	**yes**	**yes**
Analyses of data									
Point estimates of analytic sensitivity and specificity with 95% CI	NA	NA	NA	NA	NA	NA	NA	NA	NA
Sample size/power calculations addressed	no	no	no	no	no	no	no	no	no
CLINICAL VALIDITY									
Clear description of the disorder/phenotype and outcomes of interest									
Status verified for all cases	no (subgroup only/ positive cases)	no	yes	no	no (positive cases only)	no	yes (for FISH, not for NGS)	no (positive cases only)	no
Appropriate verification of controls	No controls	Unclear	Unclear	Unclear ('proven controls)'	No controls	unclear	Unclear	unclear	Yes, sequenced

Verification does not rely on index test result	No controls	unclear	unclear	unclear	No controls	unclear	unclear	unclear	no
Prevalence estimates are provided	no	no	no	yes	yes	no	yes	no	no
Adequate description of study design and test/methodology									
	no	partly	no	partly	partly	no	no	no	partly
Adequate description of the study population									
Inclusion/exclusion criteria	no	no	no	no	**yes**	no	no	no	no
Sample size, demographics	no	Only for positive samples	**yes**	**yes**	**yes**	**yes**	**yes**	no	Yes, sample size (no demographics)
Study population defined and representative of the clinical population to be tested	unclear	Unclear	unclear	**yes**	**yes**	yes	No, only BRAF neg samples, and mainly adult patients	no	No, site for major clinical (larotrectinib) trial, and major referral centre
Allele/genotype frequencies or analyte distributions known in general and subpopulations	NA	NA	NA	NA	NA	NA	NA	NA	NA
Independent blind comparison with appropriate, credible reference standard(s)									
Independent of the test	unclear	unclear	unclear	unclear	unclear	unclear	unclear	unclear	unclear
Used regardless of test results	unclear	unclear	unclear	No, reference test used only for pos samples	No, reference test used only for pos samples	unclear	All with FISH as RS, but (NGS) as RS only for pos samples	No, reference test used only for pos samples	unclear
Description of handling of indeterminate results and outliers	no	no	partly	partly	no	No	no	no	no
Blinded testing and interpretation of results	NR	NR	NR	blinded IHC interpretation	blinded IHC interpretation	NR	NR	NR	NR
Analysis of data									
Possible biases are identified and potential impact discussed	no	no	**Yes**, partly	**Yes**, partly	**Yes**, partly	**no**	**Yes**, partly	no	**Yes**, partly
Point estimates of clinical sensitivity and specificity with 95% CI	**Yes**, but no CI	no	**Yes**, ROC with CI	no	no	**Yes,** but no CI	**Yes,** but no CI	no	**Yes**, but no CI
Estimates of positive and negative predictive value	no	no	no	no	no	**yes**	no	no	**Yes**, for some cancers
Clinical utility- No data available									
Analytical validity									
-Inadequate	X	X	X	X	X	X	X	X	X
Clinical validity									
-Convincing									
-Adequate									X
-Inadequate	X	X	X	X	X	X	X	X	

CI: Confidence Interval; RS: Reference standard; NA: data/information not available; ROC: Receiver Operation Characteristics curve: 1 Bricic 2021: Normal appendix and brain tissues were used as positive controls for IHC; 2. Choi 2018: Brain ganglions and lymphocytes served as positive and negative controls, respectively; 3. As positive control tissue, a mammary analogue secretory carcinoma from the parotid gland with proven NTRK fusion was used; 4 Gatalica 2019: Placental tissue served as a positive control for PD-L1 antibodies, while cerebral cortical tissue was used for pan-Trk controls.5 Lee 2020.

Tissue from secretory carcinoma of the salivary gland with a confirmed ETV6-NTRK3 fusion and positive pan-TRK staining was applied as positive control; 6Rudzinski 2018: One case with a confirmed TPM3-NTRK1 rearrangement was used as the positive control. Non-neoplastic tissues (skin, blood vessels, inflammatory cells, renal cortical epithelium) were used as negative internal controls; 7 Salomon 2019a: These cases do not constitute internal controls, but more like clinical cases; 8 Salomon 2019: but authors also mention that the GS may also be wrong

Appendix 9 Quality of reporting- results of the STARD checklist

STARD items	Bell 2020 (43)	Bricic 2021 (44)	Choi 2018 (45)	Elfving 2021 (46)	Fu 2021 (47)	Galatica 2019 (48)	Lee 2020 (49)	Rudzinski 2018 (50)	Salomon 2020 (11)
1	Yes, In abstract	Yes, In abstract	Yes, In abstract	Yes, In abstract	No	Yes, In abstract	Yes, In abstract	Yes, In abstract	Yes, In abstract
2	no	no	no	yes	no	no	no	no	no
3	yes	yes	yes	yes	yes	yes	yes	yes	yes
4	Partly, no hypothesis	Partly, no hypothesis	Partly, no hypothesis	Partly, no hypothesis	Partly, no hypothesis	Partly, no hypothesis	Partly, no hypothesis	Partly, no hypothesis	Partly, no hypothesis
5	unclear	retrospective	retrospective	retrospective	retrospective	retrospective	retrospective	retrospective	retrospective
6	Cancer type only	Cancer type only	Cancer type only	Cancer type only	NR	yes	Cancer type only; >0.5 cm tumour	Cancer type only	NR
7	NR	register	register	Register, tumour resected	register	Register, tumour resected	register; BRAF negative	Register; not tested with FISH or NGS	Register; previously tested with NGS
8	yes	yes	yes	yes	yes	yes	yes	Place, but not time	Time but not place
9	NR	NR	NR	NR	NR	Unselected	Consecutive	NR	NR
10a	Partly (not all 3 phases)	Partly (not all 3 phases)	Partly (not all 3 phases)	Partly (not all 3 phases)	Partly (not all 3 phases)	Partly (not all 3 phases)	Partly (not all 3 phases)	Partly (not all 3 phases)	Partly (not all 3 phases)
10b	Partly	Partly	Partly	Partly	Partly	Partly	Partly	Partly	Partly
11	no	no	no	no	no	no	no	no	no
12a	yes	yes	no	yes	no	partly	partly	no	yes
12b	no	no	FISH yes (not NGS)	no	yes	FISH yes (not NGS)	FISH yes (not NGS)	no	no
13a	NR	NR	NR	NR	NR	NR	NR	NR	NR
13b	NR	NR	NR	NR	NR	NR	NR	NR	NR
14	yes	no	yes	no	no	no	no	no	no
15	NR	NR	partly	partly	NR	NR	NR	no	partly
16	NR	NR	NR	NR	NR	NR	NR	NR	NR
17	NR	NR	NR	NR	NR	NR	NR	NR	NR
18	NR	NR	NR	NR	NR	NR	NR	NR	NR
19	no	no	no	no	no	no	no	no	no
20	no	partly	yes	yes	yes	yes	yes	no	no

21a	no	no	no	yes	no	yes	yes	no	no
21b	no	no	no	no	no	no	yes	no	no
22	NR	NR	yes	NR	NR	NR	NR	NR	NR
23	yes	partly	no	no	partly	partly	partly	no	no
24	no	no	Yes (incl.CI)	no	Yes, but no CI	Yes, but no CI	no	Yes, but no CI	Yes, but no CI
25	NA	NA	NA	NA	NA	NA	NA	NA	NA
26	no	no	partly	yes	no	partly	yes	no	partly
27	yes	yes	yes	yes	yes	yes	yes	yes	yes
28	no	no	no	no	no	no	no	no	no
29	no	no	no	no	no	no	no	no	no
30	Non-commercial funders	NR	Non-commercial funders	Non-commercial funders	Commercial funders, role NR	Non-commercial funders	Non-commercial funders	NR	Commercial funders, role NR

STARD checklist items:
TITLE or ABSTRACT
1 Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, PPV, NPV or AUC)
ABSRACT
2 Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION
3 Scientific and clinical background, including the intended use and clinical role of the index test
4 Study objectives and hypotheses
METHODS
Study design
5 Whether data collection was planned before the index test and reference standard were performed (prospective) or after (retrospective study)
Participants
6 Eligibility criteria
7 On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
8 Where and when potentially eligible participants were identified (setting, location and dates)
9 Whether participants formed a consecutive, random or convenience series
Test methods
10a Index test, in sufficient detail to allow replication. Note: All three phases should have been described: pre-analytical, analytical, post-analytical.
10b Reference standard, in sufficient detail to allow replication
11 Rationale for choosing the reference standard (if alternatives exist)
12a Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
12b Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
13a Whether clinical information and reference standard results were available to the performers/readers of the index test
13b Whether clinical information and index test results were available to the assessors of the reference standard
Analysis
14 Methods for estimating or comparing measures of diagnostic accuracy
15 How indeterminate index test or reference standard results were handled
16 How missing data on the index test and reference standard were handled
17 Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
18 Intended sample size and how it was determined
RESULTS
Participants
19 Flow of participants, using a diagram
20 Baseline demographic and clinical characteristics of participants
21a Distribution of severity of disease in those with the target condition
21b Distribution of alternative diagnoses in those without the target condition
22 Time interval and any clinical interventions between index test and reference standard
Test results
23 Cross tabulation of the index test results (or their distribution) by the results of the reference standard
24 Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
25 Any adverse events from performing the index test or the reference standard
DISCUSSION
26 Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
27 Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION
28 Registration number and name of registry
29 Where the full study protocol can be accessed
30 Sources of funding and other support; role of funders

Appendix 10 Funding, role of funders and conflicts of interest

Author Year	Financial support	Role of funder	Conflict of interest	Comment
Original papers (N=9)				
Bell 2020 (43)	Supported in part by the Head and Neck SPORE Program Grant P50CA097007, The Kenneth D. Muller Professorship (Adel El-Naggar, MD, PhD)	NR	The authors declare no conflict of interest.	Non-commercial funding.
Bricic 2021 (44)	NR	NR	The authors declare no conflict of interest	Unclear funding, but authors declare no conflicts of interest.
Choi 2018 (45)	Supported by the National Research Foundation of Korea (NRF) funded by the Korean government (MSIP) (grant no. 2014R1A2A1A11052217); and by a grant from the National Research and Development Program for Cancer Control, Ministry for Health, Welfare and Family Affairs, Republic of Korea (grant no. 0920050). The biospecimens for this study were generously provided by the Pusan National University Hospital and National Biobank of Korea, which is supported by the Ministry of Health, Welfare, and Family Affairs	NR	The authors declare no conflict of interest	Non-commercial funding.
Elfving 2021 (46)	Supported in part by the Swedish Cancer Society, the Selanders Foundation Uppsala and the Lions Cancer Foundation Uppsala	NR	The authors declare no conflict of interest.	Non-commercial funding
Fu 2021 (47)	Supported by grants from the Chinese National Science Foundation (81802394 to XP) and Fundamental Research Funds for the Central Universities (021414380408 to XP).	NR	The authors declare no conflict of interest.	Non-commercial funding
Galatica 2019 (48)	Caris Life Sciences conducted this study.	NA (industry)	Zoran Gatalica, Joanne Xiu, and Jeffrey Swensen are all employees of the Caris Life Sciences. Semir Vranic has received honoraria from Caris Life Sciences.	
Lee 2020 (49)	Supported by the research grants from Taipei Veterans General Hospital (Grant No.: V109B-029), Taipei Veterans General Hospital-National Yang-Ming University Excellent Physician Scientists Cultivation Program (Grant No.: 109-V-B-002 and 109-V-B-003), and Taipei Institute of Pathology (Grant No.: TIP-108-004).	NR	The authors declare no conflict of interest.	Non-commercial funding
Rudzinski 2018 (50)	NR	-	The authors declare no significant relationships with, or financial interest in, any commercial companies pertaining to this article	Unclear funding, but authors declare no conflicting financial interests.
Salomon 2020 (11)	This study was funded by the National Cancer Institute (NCI) under the MSK Cancer Center Support Grant/Core Grant (P30 CA008748) and a research grant from Bayer AG.	NR	ER is supported by internal Memorial Sloan Kettering funding and Memorial Sloan Kettering Clinical Scholars grant 2T32CA009512-29A1. DMH is in a consulting or advisory role for Chugai Pharma, CytomX Therapeutics, Boehringer Ingelheim, AstraZenica, Pfizer, Bayer, and Genentech, and has received research funding from AstraZenica, Puma Biotechnology, Loxo, and Bayer. AD reports honoraria from Ignyta/ Genentech/Roche, Loxo/Bayer/Lilly, Takeda/Ariad/Millenium, TP Therapeutics, AstraZeneca, Pfizer, Blueprint Medicines, Helsinn, Beigene, BergenBio, Hengrui Therapeutics, Exelixis, Tyra Biosciences, Verastem, MORE Health, Merck,	Project partly commercially funded. Many of the authors have received personal financial support.

			Puma, Medscape, OncLive, PeerVoice, Physicians Education Resources, Targeted Oncology, and Research to Practice, research funding from Pfizer, Exelixis, GlaxoSmithKlein, Teva, Taiho, PharmaMar, and Foundation Medicine, and royalties from Wolters Kluwer. JFH has received honoraria from Axiom Healthcare Strategies, Cor2Ed, and Medscape, as well as research funding from Bayer. The remaining authors declare that they have no conflict of interest	
Narrative reviews, and expert opinion (ESMO) papers (N=7)				
Gambella 2020 (51)	Supported by the Rete Oncologica del Piemonte e della Valle d'Aosta and received funding specifically dedicated to the Department of Medical Sciences, University of Turin from Italian Ministry for Education, University and Research (Ministero dell'Istruzione, dell'Università e della Ricerca-MIUR) under the programme "Dipartimenti di Eccellenza 2018 – 2022", Project n◦ D15D18000410001.	The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.	The authors declare no conflict of interest	Non-commercial funding
Hsiao 2019 (52)	Supported by Loxo Oncology, Inc., and Bayer in the form of funding for medical writing and open access publication.	.NR	S.J.H. has received honoraria from BMS and Roche; A.N.S. is an employee of and owns stock in Loxo Oncology, Inc; D.L.A. has consultancy roles with Bayer Oncology, Bristol-Meyers Squibb, AstraZeneca, Genentech, and AbbVie and has received research funding from Genentech	Project commercially funded. Many of the authors received financial support from commercial suppliers.
Marchia 2019 (18)	Funded by European Society for Medical Oncology (no grant number applies).	NR	CM has received personal/consultancy fees from Axiom Healthcare Strategies, Cor2Ed, Bayer and Daiichi-Sankyo. MS received research funds from Puma Biotechnology, Daiichi-Sankyo, Immuno medics and Menarini Ricerche, honoraria from ADC Pharma and Menarini Ricerche and is a cofounder of Medendi Medical Travel. AJI is a founder and equity holder inArcherDx, has received research funds from Blueprint Medicines,and is a consultant for Chugai, DebioPharm, Constellation, Rocheand Pfizer. ML has received advisory board compensation from Boehringer Ingelheim, AstraZeneca, Bristol-Myers Squibb, Takeda and Bayer, clinical research support from LOXO Oncology, and pre-clinical research support from Helsinn Healthcare. JFH has received honoraria from Medscape, Axiom Biotechnologies andCor2Ed, as well as research funding from Bayer. FL-R has received research funding from Thermo Fisher and Roche, as well as honoraria from Bayer, Thermo Fisher and Roche. JSR-F reports personal/consultancy fees from Volition Rx, Page.AI, Goldman Sachs, Grail, Ventana Medical Systems, Roche and Genentech. All remaining authors have declared no conflicts of interest.	No-commercial funding Many of the authors received financial support from commercial suppliers.

Perault-Llorca 2019 (19)	Supported by Bayer Healthcare Pharmaceuticals and Loxo Oncology according to Good Publication Practice guidelines (http://annals.org/aim/ article/2424869/good-publication-practice-communicating-company-sponsored medical-research-gpp3).	NR	FP-L has participated in advisory boards for Bayer, Roche, Illumina and Nanostring, and been involved in studies sponsored by Bayer. ERR has had a role as an expert consultant, participated in a meeting and participated in an advisory board for Bayer Healthcare Pharmaceuticals. ARS has had a role as an expert consultant for Merck US, Bristol Meyers Squibb and Bayer Healthcare Pharmaceuticals; participated in meetings for Merck US, Bristol Meyers Squibb and Bayer Healthcare Pharmaceuticals; participated in advisory boards for Merck US, Bristol Meyers Squibb and Bayer Healthcare Pharmaceuticals; and received honoraria from Amgen.	Project commercially funded. Many of the authors have also received money from commercial suppliers
Salomon 2019b (55)	NCI under the Cancer Center Support Grant/Core Grant (P30 CA008748) awarded to Memorial Sloan Kettering Cancer Center	NR	J.F. Hechtman reports receiving commercial research grant from Bayer and reports receiving honoraria from Medscape, Cor2Ed, and Axiom Healthcare Strategies. No potential conflicts of interest were disclosed by the other author.	Non-commercial funding. One of the authors received financial support from commercial suppliers
Salomon 2019c (53)	This paper was published as part of a supplement financially supported by Bayer AG and Loxo Oncology, Inc., a wholly owned subsidiary of Eli Lilly and Company.	NR	JFH has received honoraria from WebMD, Axiom Healthcare Strategies and Cor2Ed, and research funding from Loxo Oncology, Bayer and Boehringer Ingelheim. ML has received research funding and advisory board compensation from Loxo Oncology and advisory board compensation from Bayer. All remaining authors have declared no conflicts of interest.	Project commercially funded. Many of the authors received financial support from commercial suppliers
Wong 2019 (54)	This work was partly supported by funds from CIHR Foundation grant FDN-143280 (to PHS), Brain Care BC (to SY), and the BC Cancer Foundation (to PHS).	NR	Dr. Wong declares no conflicts of interest. Dr. Yip declares consultation fees from Bayer and Pfizer for his participation in advisory boards and travel expense reimbursement from Roche/ Foundation Medicine. Dr. Sorensen declares that he is an advisor for Bayer Pharmaceuticals but holds no financial interest in the company.	Non-commercial funding. Two of the authors received financial support from commercial suppliers.

Appendix 11 Estimated cost for NTRK and ROS1 testing with different methods, St. Olav's University Hospital and Oslo University Hospital

It is notable that based on the feedback we received from St. Olav's University Hospital and Oslo University Hospital, NGS is currently used at these hospitals for the detection of NTRK gene fusions (personal communication).

IHC

Estimated cost for testing with IHC, St Olav's University Hospital and Oslo University Hospital

	St. Olav's University Hospital		**Oslo University Hospital**
	1 patient (NOK)	**10 patients (NOK)**	**1 patient (NOK)**
Reagent cost	408	408	250
Personal cost	564 (bioengineer: 384.62, pathologist: 179.49)	269 (bioengineer: 89.74, pathologist: 179.49)	300 (bioengineering: 210, Pathologist: 90)
Sum	**972**	**677**	**550**

FISH

Estimated cost for testing with FISH, St Olav's University Hospital and Oslo University Hospital

	St. Olav's University Hospital		**Oslo University Hospital**
	1 patient (NOK)	**10 patients (NOK)**	**1 patient (NOK)**
Reagent cost	1,500	1,500	1,094
Personal cost	930 (bioengineer: 810, pathologist: 120)	364 (bioengineer: 244, pathologist: 120)	1,967 (bioengineering: 1,280, Pathologist: 687)

Sum	**2,430**	**1,864**	**3,061**

RT-PCR

*Estimated cost for testing with RT-PCR, St Olav's University Hospital**

	St. Olav's University Hospital	
	1 patient (NOK)	**10 patients (NOK)**
Reagent cost	1,200	1,200
Personal cost	930 (bioengineer: 810, pathologist: 120)	364 (bioengineer: 244, pathologist: 120)
Sum	**3,113**	**1,419**

** RT-PCR is not used for NTRK and ROS1 testing in Oslo University Hospital*

NGS

Estimated cost for testing with NGS, St Olav's University Hospital and Oslo University Hospital

	St. Olav's University Hospital*		**Oslo University Hospital** (NOK)**
	1 patient (NOK)	**10 patients (NOK)**	
Reagent cost	17,580	2,580	7,168
Personal cost	1,462 (bioengineer: 948.70, molecular biologist: 333.33 pathologist: 179.49)	705 (bioengineer: 192.30, molecular biologist: 333.33 pathologist: 179.49)	3,800 (bioengineering:3,000, Pathologist: 800)
Sum	**19,042**	**3,285**	**10,968**

**Ion Torrent S5 (personal communication by Liv Solvår Nymark, St. Olav's University Hospital)*

***Ion Torrent S5, Oncomine Childhood Cancer Research Assay (personal communication by Martin Andreas Furu, Oslo University Hospital)*

Activity log

Suggestion submitted	10.05.2019 (ID2019_029) 24.10.2019 (ID2019_119)
HTA commissioned	26.08.2019 (ID2019_029) 16.12.2019 (ID2019_119)
External clinical experts pointed out by RHA	20.08.2020
Work paused due to prioritization of ROS1 report	01.09.2020
Updated search for literature	03.05.2021
Report sent for expert review	07.12.2021
Internal approval at NIPH	22.12.2021
Report submitted	23.12.2021

Published by the Norwegian Institute of Public Health
Januar 2022
P.O.B 4404 Nydalen
NO-0403 Oslo
Phone: + 47-21 07 70 00
The report can be downloaded as pdf
at www.fhi.no/en/publ/