Using artificial INtelligence to Support Informed Decision-making (INSIDE) on BRAF testing and mutation

Authors: Scott Kopetz,[1] Christopher H Lieu,[2] Orion Penner, Bob Schijvenaars,[3] Jennifer Ghith,[4] Jennifer Webster,[4] Georgina Long,[5,6] Richard Scolyer[5,7,8,9] (author list order TBD)

[1]Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[2]University of Colorado Denver School of Medicine, Aurora, Colorado, USA

[3]Digital Science, London, UK

[4]Pfizer Inc., New York, NY, USA

[5]Melanoma Institute of Australia, The University of Sydney, Sydney, NSW, Australia

[6]Royal North Shore and Mater Hospitals, North Sydney, Sydney, NSW, Australia

[7]Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia

[8]Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital and NSW Health Pathology, Sydney, New South Wales, Australia

[9]Charles Perkins Centre, The University of Sydney, Sydney, New South Wales, Australia

Introduction

- In clinical oncology, precision medicine largely relies on the identification of oncogenic mutations. Crucially, to quantify the number of people who could potentially benefit from precision medicine, epidemiologists rely on testing and mutation rates. In turn, this number influences the size of the investment into the development of novel diagnostics and precision therapeutics by healthcare providers and healthcare systems, and, ultimately, the likelihood that these technologies will be used [1-3]. It is therefore important to ensure that both testing and mutation rates, particularly for biomarkers used to inform treatment decisions, are reported, and reported clearly and consistently, within the scientific literature.
- Although a number of guidelines, including those by the European Society for Medical Oncology (ESMO) and the US National Comprehensive Cancer Network (NCCN), recommend routine testing of oncogenic driver mutations across a range of cancers (e.g., advanced non-small cell lung carcinoma [NSCLC] [4, 5]), they do not provide guidance on best practices for clear and consistent reporting of these data in published literature.
- Joint guidelines such as those from the American College of Medical Genetics (AMP)/American Society of Clinical Oncology (ASCO)/College of American Pathologists (CAP) provide some guidance on oncogene somatic mutation data sharing between testing laboratories and larger national/international databases

[6], but again, no guidance is provided on reporting these data in published literature.
    - Furthermore, the literature describing adherence to the testing recommendations contained in these guidelines is sparse.
- Precision oncology is predicated on the observation that treatments can be made more effective when matched to individuals based on both their germline genetic profile and that of the tumour, which may vary from person to person and change over time.
    - It is a rapidly evolving field, necessitating that clinicians frequently review the most recent literature; a challenging prospect in the face of high publication volume and velocity (Figure X)

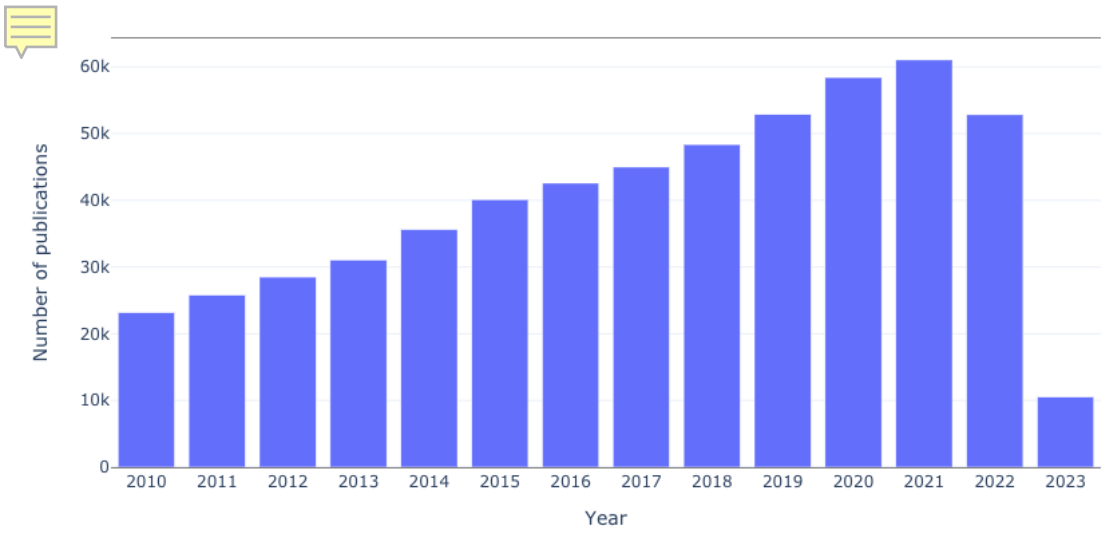

Figure X Estimate of the annual number of precision oncology-related publications. The PubMed search query and results used are available in Supplementary File 1.

- Inadequate and/or unclear reporting of testing and mutation rates can have negative implications for researchers, industry, and healthcare providers, which may ultimately result in poorer clinical outcomes for patients.
    - For researchers and industry, ambiguous reporting of testing and mutation rates can lead to inappropriately powered clinical trials; should a limited number of patients harbour a specific mutation, then trial recruitment may be challenging and compromise study power [7].
    - If reported rates are inflated, then identification of patients with mutations that may benefit from targeted therapy may be more time-consuming and costly than anticipated owing to the increased screening burden [7].
    - If there is reason to believe that a sub-population of patients may benefit more than another from precision medicine, allocation of patients in a clinical trial to a potentially inferior treatment would be unethical [8].
    - Furthermore, samples taken for research may not be representative of the general population [9]. For example, research may be conducted on people of a given ethnicity who are more likely to have certain mutations than people of another ethnicity. An instance of this phenomenon can be found with *EGFR* mutations in patients with non-small-cell lung adenocarcinoma: such mutations are two and a half times more likely in patients of Asian ethnicity than in patients of Caucasian ethnicity [10]. On a finer scale, research samples may also not be representative of the entire tumour cell population [9].

  - Focus on certain therapeutics based on inaccurate biomarker epidemiology may be skewed, leading to inefficient use of resources in targeting a population of patients far smaller than anticipated. Physicians may therefore order tests that are unlikely to be warranted or researchers invest efforts on drug targets that could be better spent elsewhere [11].
  - Should a personalized medicine approach be warranted, the medicine and the associated technologies of genomic testing will likely be expensive, so economic arguments may hinder access in both developed and developing worlds and exacerbate gaps between patients based on wealth [9].
  - For healthcare professionals, flawed epidemiological data can potentially lead to the under adoption and inappropriate use of genomic testing for well-described oncogenic driver mutations, that in turn may negatively impact efficient treatment decisions [11].
  - For patients, inflated rates may lead to unnecessary tests and lost time, critically important when treating cancer. Conversely, deflated rates may result in reduced testing and missed treatment opportunities.

- Real-world data can be helpful in answering questions that can be less easily addressed by prospective clinical trials, for example, longer-term outcomes or those of patients not meeting clinical trial eligibility criteria.
- Genomic data are now an important source of real-world data and may improve our understanding of patients with relatively uncommon mutations for whom entry into a clinical trial is unlikely, to obtain information on prognosis or clarify the impact of treatment sequencing decisions [12].
- There are many real-world studies that report testing and mutation rates. The sheer volume of such data, and the inconsistency in use of genetic tests and how they are reported, makes manual aggregation of data across the literature time-consuming and difficult to perform [13].
- An approach that leverages artificial intelligence (AI) to assist in literature searching and data extraction presents a potential solution to this problem; automating this process may offer efficiencies compared with manual review and improved ability to identify the most relevant information [13-15].
  - There is growing evidence to support the applicability [16] and usefulness [17] of this process [18].
  - Literature reviews may also benefit from AI for natural language generation, using tools like Chat Generative Pre-trained Transformer (ChatGPT) [19] or Generative Pre-trained Transformer 3 (GPT-3) [20]. However, these AI tools can plagiarise content by not attributing their sources [21, 22], cannot explain their reasoning [23] and sometimes display over-confidence in their incorrect responses [24]. Tools available in the future may lift some of these restrictions (for example the AI tool Sparrow cites its sources [25]); at present, applications not reliant on unknown sources can be considered, such as summarizing the content of a specific set of publications.
  - Models like bidirectional encoder representations from transformers for biomedical text mining (BioBERT) [26] are ideally suited to this task because they combine the power of transformer models with a biomedical specialisation, acquired through pretraining on many sentences in the biomedical domain.
  - The efficiency gains afforded by AI automation could reduce the costs of the comprehensive collection of biomarker data from the literature [27] and permit literature reviews for diseases with lower patient numbers than in oncology,

promoting health equity. This would facilitate the identification of knowledge gaps, an initial step in the medical research and development process.

- This study specifically examines reporting of testing and mutation rates for *BRAF*, an important oncogenic driver that, when present, is indicative of a poor prognosis in patients with metastatic colorectal carcinoma (mCRC), advanced non-small cell lung cancer (aNSCLC) or advanced cutaneous melanoma (aMelanoma) [28-31].
    - Using *BRAF* as the pilot allows comparison among multiple cancer types. Describing the *BRAF* testing landscape allows *BRAF* to be used as a proxy for other genes.
    - *BRAF* testing is critical for personalised treatment decisions [31, 32], although testing technologies can carry the risk of false negatives [32].
    - *BRAF* testing rates are highly variable globally [33] and guidelines do not mandate automatic, pathologist-initiated testing [34]. In addition, some evidence points to inequity in testing, depending on regional, socio-economic and ethnic factors [35-37].
    - Thus, many patients may not benefit from a growing range of efficacious, licensed precision therapies against this key oncogenic driver [38].

Aims

We aimed to use AI to identify and extract *BRAF* testing and mutation rates from the literature including patients with mCRC, aNSCLC or aMelanoma. By assessing the quality of testing and mutation rate reporting with AI, our broader objectives are to (1) test the capacity of AI to aid in summarizing the literature; (2) propose appropriate standards of rate reporting if required; (3) contribute to the research and development of precision medicine; (4) improve patient access to the most effective treatments.

Methods

Definitions

BRAF mutation

- We focused on BRAF mutation, defined as a change in one of the amino acids that compose the BRAF protein. Chromosomal abnormalities or copy number variants were not considered, although these can also be associated with disease [39, 40].
- One of the mutations considered was base substitutions, which we identified by the following code:

$$WPosM$$

in which W is the wild-type (non-mutated) amino acid at position Pos within the BRAF protein and M is the mutant amino acid that replaced W at that position. For example, one of the most common BRAF mutations is coded as V600E, which means that the amino acid glutamic acid (E) replaced valine (V) at position 600 in the protein.

- We also considered two other genetic variants: first, an amino acid deletion-insertion, coded as 'V600-K601delinsE', which means that the amino acids at position 600 (valine; V) and at position 601 (lysine; K) in the protein were replaced by glutamic acid (E); second, an amino-acid duplication, coded as 'T599dup', which signifies that the amino acid at position 599 (threonine; T) was copied and inserted next to the original towards the C-terminus of BRAF.

Mutation and testing rates

- We identified publications that included patients who were diagnosed with mCRC, aNSCLC or aMelanoma and who were enrolled in a study (observational or interventional). From these publications, we extracted the testing and mutation rates in these patients.
- We defined the testing rate as the proportion of included patients who were tested for a *BRAF* mutation during normal clinical care. Formally,

$$testing\ rate = \frac{n_{\text{prior}}}{N},$$

in which n_{prior} is the number of patients who were tested prior to their inclusion in any identified study and N is the total number of included patients.

- We defined the mutation rate as the proportion of positive patients among those tested for a *BRAF* mutation as part of an identified study. Therefore,

$$mutation\ rate = \frac{n_+}{n},$$

in which n_+ is the number of patients who tested positive for a *BRAF* mutation as part of the study, and n is the number of patients who were tested for a *BRAF* mutation as part of the study.
- We define the mutation subtype rate as the proportion of patients with a *BRAF* mutation in a specific set, among the patients tested for a *BRAF* mutation during the study. That is,

$$mutation\ subtype\ rate = \frac{n_{M_+}}{n},$$

in which n_{M_+} is the number of patients who tested positive for a *BRAF* mutation in the set, M, as part of the study. For example, this set may be restricted to the V600E mutation, in which case the mutation subtype rate is just the rate of V600E mutation among those tested, or the set may include both the V600E and V600K mutations, in which case the mutation subtype rate is the rate of V600E or V600K mutation among those tested.

Natural language processing

Figure X Visual summary of the methodology used for natural language process and statistical analysis

- Our methodology to train, validate and test AI for natural language processing is summarised in Figure X.

Step 1: corpus selection with literature searches

- We searched the literature to find publications that would potentially contain either a testing rate or a mutation rate. The search was conducted in the Dimensions database, retrieved English-language documents published between 2017 and 2021, and included general oncology terms such as “cancer”, “neoplasms”, “melanoma”, or “carcinoma”. In addition, we searched for phrases specific to BRAF testing and mutations rates.

- To identify publications including testing rates, search phrases were generated using the following template

$$"reception\ target\ test"$$

in which $reception$ is a verb phrase that indicates the testing procedure, $target$ is an adjectival phrase that describes the aim of the procedure, and $test$ is a noun phrase that refers to the outcome of the procedure. For example, “received ($reception$) BRAF ($target$) testing ($test$)” and “underwent ($reception$) genetic ($target$) profiling ($test$)”.

Search phrases for publications including mutation rates were generated using the following template

$$"protein\ mutation\ observation"$$

in which $protein$ is a phrase that refers to a mutated protein, $mutation$ is a phrase that denotes a mutation, and $observation$ is a phrase that is related to the assessment of this mutation. For example, “BRAF ($protein$) mutations ($mutation$) observed ($observation$)” and “V600K ($protein$) genetic alteration ($mutation$) evidence ($observation$)”. This templated generation process resulted in 273 search phrases for testing rates and 280 for mutation rates (Supplementary Table 1).

- A final search string found documents that contained one general oncology search term, and one mutation- or testing-related search phrase (e.g. ‘(“cancer” OR … OR “melanoma”) AND (“received BRAF testing” OR … OR “underwent genetic profiling”)’). This process resulted in the creation of two search strings, one for identifying publications that include testing rates, and the other for those including mutation rates. Following these searches, the list of documents was filtered to exclude certain document types, such as “books” and “monographs”, resulting in 16 617 retained documents.

Step 2: filter for cancer type using AI

- Using machine learning, the documents retrieved in step 1 were further classified as concerning one of three types of cancer (lung, colorectal, melanoma), or none of these. These three types are referred to as ‘target’ cancer types; other cancer types are called ‘non-target’.

- Three BioBERT models (one for each target cancer type) were fine-tuned to perform this classification, each using a linear layer for decoding the abstract representations (‘contextual embeddings’) of the documents. Each model was fine-tuned with 10 000 publications (training set) using pre-existing Medical Subject Headings (MeSH) labels found on PubMed, which indicated whether these publications were related to the target cancer types or to the non-target cancer types. Each training set consisted of 25% ‘positive examples’ (documents tagged with the target cancer type) and 75% of ‘negative examples’ (documents tagged with non-target cancer types). Fine-tuning minimized a metric that quantifies the classification error, called cross-entropy loss, and models were not penalized for their complexity (no regularization).

- Using a testing set of 10 000 publications for each model, we derived performance measurements using the precision and recall metrics, defined as

$$precision = \frac{correctly\ predicted\ positives}{predicted\ positives},$$

$$recall = \frac{correctly\ predicted\ positives}{actual\ positives},$$

in which $correctly\ predicted\ positives$ is the count of documents correctly predicted as being related to the target cancer type, $predicted\ positives$ is the sum of both correct and incorrect predictions, and $actual\ positives$ is the count of documents with the target cancer type in the examples tagged with MeSH terms. The performance measurements are presented in Table X.

Table X Performance of the AI models for cancer type classification in the testing set

Cancer type	**Precision (%)**	**Recall (%)**
Lung cancer	97.8	94.3
Colorectal cancer	98.8	96.0
Melanoma	98.5	93.7

- As a result of this step, 7832 documents were excluded, which means that 8785 documents were retained.

Step 3: sentence classification using numerical motifs and AI

Using punctuation to indicate sentence boundaries, we extracted the sentences from the remaining documents and stored them in a table that enables fast retrieval (lookup table). Terms that described 11 patterns of characters (regular expressions) were used to identify sentences with a numerical motif (Supplementary Table 2). This further reduced the corpus to 7958 documents containing 365 683 sentences.

We trained a BioBERT model with a linear decoding layer to predict the probabilities that a sentence contained a mutation rate, a testing rate, or no rate. A data set for model training and testing was obtained by manually reviewing 2244 sentences to determine the presence and type of rate; as a result, 1570 sentences (70%) were used for training and 674 (30%) were used for testing. Cross-entropy loss was minimized during training, and no regularization was applied. We used the probabilities provided by BioBERT to classify the probable presence of rate data and rate type, setting thresholds to obtain 75.0% recall in the testing set. The performance measurements are presented in Table X.

Table X Performance of the AI models for sentence classification in the testing set

Rate type	**Precision (%)**	**Recall (%)**
Testing rate	74.3	75.0
Mutation rate	53.3	75.0
No rate	98.9	75.0

Because BioBert provided probabilities that a sentence contained a mutation rate, a testing rate or no rate independently, the process may result in conflicts, with a sentence possibly classified as both containing and not containing a rate. In this case, we considered that the sentence was predicted to contain a rate, because no-rate sentences were later discarded and we aimed to keep these ambiguous cases for the later stages of the process. When a sentence was classified as both containing a mutation and a testing rate, the sentence received both provisional classifications.

The classification was used to filter sentences down to 1687 testing rate sentences and 14 854 mutation rate sentences, contained within 963 documents.

Step 4: manual annotation using a consensus methodology

The documents shortlisted in the previous step were shared with expert analysts to identify mutation rates, testing rates (or the population sizes needed to compute them) and covariates, including the country where the study took place, the cancer type and the specific *BRAF* mutation, the type of the study (e.g. “observational retrospective/prospective” or “clinical trials”), the characteristics of the treating hospital and the demographics of the patients (gender and race distribution), when available. These entities were identified in the full text of the publications. The full list of identified entity types in publications containing testing rates is available in Supplementary Table 3 and the list of identified types in publications with mutation rates is given in Supplementary Table 4.

- Mutation and testing rates were independently annotated by three experts, and discrepancies were addressed using a consensus methodology. Agreement on a document was defined as being achieved if all three experts annotated the same mutation or testing rates, or if they did not annotate any rate within the document.

    - During the initial round of manual annotation, experts discarded 12% (117/963) documents that they identified to be discussing a non-target cancer type. As a result of this round, the three experts agreed on the identification of X% (n/N) of testing rates and of only 30% (398/1333) of mutation rates, resulting in agreement for X% (n/N) of documents with testing rates and for 38% (320/848) of documents with mutation rates.

    - In the first consensus round, disagreement was resolved by showing to the reviewers the anonymized annotations of the other reviewers, enabling agreement for a further X% (n/N) of testing rates and 25% (328/1333) of mutation rates. Note that for mutation rates, this round only examined instances for which two experts agreed and one disagreed, to reduce annotation time given the high disagreement in the previous round.

    - In the second consensus round, analysts jointly analysed documents to reach a consensus. With this method, they annotated the remaining X discrepant testing rates and an additional 399 discrepant mutation rates. These mutation rates were taken from studies that included more than 135 patients, so that mutation rates for 80% of patients in all included publications were identified. These annotation and consensus steps are summarised in Supplementary Figure 1 for testing rates and in Supplementary Figure 2 for mutation rates.

- Covariates were independently annotated by 3 experts and disagreements were resolved by choosing the annotation of the majority.

- We call *testing-rate data set* the entities extracted in publications containing testing rates and *mutation-rate data set* the entities extracted in publications with mutation rates.

Figure X Construction of the testing- and mutation-rate data sets.

Interactive data platform

Extracted mutation and testing rates can be explored using an interactive platform accessible at https://braf.dimensions.ai/. The mutation and testing rates are presented in two separate tabs and can be filtered based on a choice of covariates. For example, users may choose to only display rates that were calculated with patients with aMelanoma and that were extracted from studies published between 2019 and 2021.

Using an interactive map, users can visualize the arithmetic mean of rates for each country , and each extracted rate can be visualized on a forest plot. The forest plot provides confidence intervals for each rate and a link to the originating study. We summarised these rates with a pooled rate, which is computed as the arithmetic mean of the rates weighted by a random-effects model.[1] We also provide a *p*-value obtained from a statistical test of heterogeneity (chi-squared test), which helps to determine whether the distributions of the rates are sufficiently similar to warrant their combination into the pooled rate. Platform users have the option of downloading the data for further analysis.

Detail on the procedure and equations that we used are given in Neyeloff *et al.* (2012) [42], except for the following computation. We measured the sampling error of each individual rate by computing its standard error, SE, which is derived from the standard deviation of a Bernoulli distribution [43], as

$$SE = \sqrt{\frac{r(1-r)}{n}},$$

in which r is a rate extracted from a study and n is the size of sample used to compute that rate.

Results

- A sample of the extracted rates is presented below. Figure X presents the extracted testing rates in patients affected by aMelanoma, and Figure X presents the extracted mutation rates for patients with the V600E mutation and aMelanoma.

Figure X Forest plot of the extracted testing rates for patients with aMelanoma
aMelanoma, advanced cutaneous melanoma.

Figure X Forest plot of the extracted mutation rates for patients with the V600E mutation and aMelanoma
aMelanoma, advanced cutaneous melanoma.

Interactive platform

- The interactive platform (accessible at https://braf.dimensions.ai/) offers a reliable source of information that summarizes the vast literature reporting testing and mutation rates.
- Rates are intuitively presented using a world map and a forest plot (Figure X).
- The interactivity of the platform enables users to explore the data based on their personal queries. For example, the user may choose to display rates for one of 32 mutation subtypes, (e.g. V600E).
- Interpretation of the results is facilitated by accompanying text, as well as a suitable cautionary message in case of possible misinterpretation, based on the *p*-value.

[1] These weights incorporate the variance of the distribution of true rates, which we estimate with the DerSimonian-Laird estimator. [41]

- This platform offers a rapid alternative or a prelude to a potentially lengthy article search on a scientific search engine (e.g. PubMed). Such searches typically require complex Boolean search strings, which need to be refined iteratively.

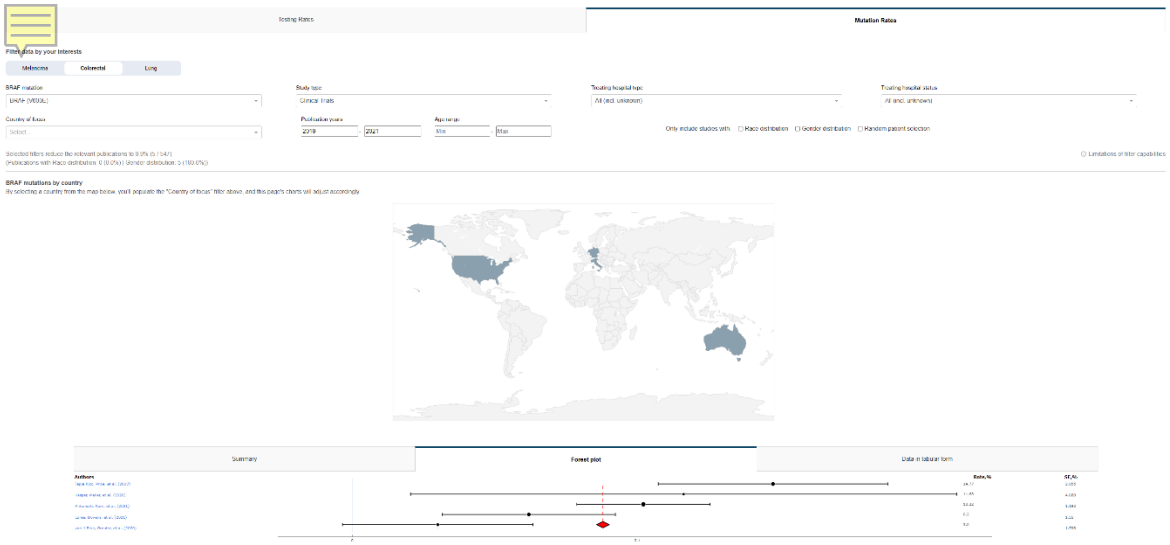

Figure X View of elements of the interactive platform: filters, summary statistics, forest plot and source publications. The platform can be accessed here: https://braf.dimensions.ai/ [website accessed on 2 March 2023].

Lexicon analysis

The manual annotation of rates by expert analysts revealed moderate disagreement (X% of cases) for testing rates and considerable disagreement (70% of cases) for mutation rates. Such high rates of disagreement may stem from unclear or non-standard reporting of rates. We present a sample of testing and mutations rates for which there was agreement in Supplementary Table 5, and a sample for which there was disagreement in Supplementary Table 6.

Discussion and expert recommendations

- Summary of reporting of *BRAF* testing and mutation rates in the literature (e.g., frequency/heterogeneity)
- Interpretation and implications of the results of the meta-analysis (main focus of the discussion).
- Need for clear and consistent reporting
    - Manual literature review demonstrated a paucity of well-reported testing and mutation rate data
    - AI was proposed as a potential solution and was used to handle large volumes of testing and mutation data, but the success of this method was limited by an heterogenous and opaque reporting lexicon of testing and mutation rates
    - In addition to this, although genetic testing is recommended for numerous cancers, there are no recommendations or guidelines regarding methods for reporting genetic testing and mutation rates in clinical trials and the scientific literature [28]
    - Deficiencies in the reporting of the testing and mutation rates of oncogenes as well as in the reporting of demographic data were identified, which potentially

negatively impacts all stakeholders and raises questions of equity of access to clinical trials, biomarker testing and precision medicine.
    - Solution: There is a need for more consistent and clear reporting of testing and reporting rates of biomarkers. Authors of studies should ensure this information is collected and included such that it is clear and easy to find, and editors of journals should consider standardising the presentation of this information in papers submitted to their journal
- Validity of AI
    - Description of whether this method worked
    - Description of how well each AI-driven method performed compared with manual curation
- Potential of AI
    - This AI-driven tool and its framework could potentially be broadly extended to collect data on other biomarkers
    - As demonstrated in this study, AI was able to search and identify articles that potentially contained information on *BRAF* testing and mutation rates with accuracy and with greater efficiency than manual literature searching and data extraction
    - This reinforces the potential for AI tools to allow for equitable collection of biomarker data in different diseases, allowing researchers to more easily identify areas of patient need and gaps in the literature, as well as speed up their research and development processes
    - Should our recommendations be implemented then it is highly likely that AI will then be able to support in the way that we initially hoped. That is to scan vast amounts of literature, identify relevant passages and then extract the numerical values that report testing and mutation rates.
        - Importantly, while we focused on *BRAF*, our results and its implications are widely applicable, as inadequate reporting of *BRAF* mutations is likely to be representative of reporting among all genomic and other biomarkers.
        - Our text processing pipeline can be easily adapted to similar use-cases.
- Limitations of study
    - There are different types of genetic tests (such as DNA/RNA or protein level, somatic or germline mutations), all of which have different sensitivities. This study, and the AI tool, did not differentiate between different types of tests. Moreover, many labs are encouraging use of simultaneous somatic and germline testing.
    - We did not search for mutations reported in the format recommended by the Human Genome Variation Society (e.g. p.(Arg54Ser)), but favoured the more informal, but commonly used format described in the Methods section.
    - We did not identify variation in the number of copies of *BRAF* nor non-protein-coding mutations, although they may worsen clinical outcomes [39, 44].
    - Our approach using artificial intelligence to reduce the burden on manual annotation may have resulted in missing more rates than a fully manual approach. In particular, our sentence classification models were expected to miss 25% of rates (75% recall), a choice made to maintain precision.
    - Limitations regarding the fact that this is not a randomized study measuring the mutation and testing rates, but rather an interactive visualization of rates published in the literature.
    - Limitations regarding the statistical analysis (this is not a meta-analysis, e.g. no sensitivity analysis).

- Call to action i.e., better reporting and why this is needed
    - Genetic testing and precision medicine positively influences patient care in oncology, and will continue to do so in the future [45, 46]
    - As genetic testing becomes more prevalent, it is important that accurate epidemiological data regarding testing and mutation rates are reported, and that this is communicated in a way that is understandable to clinicians.
        - This will ensure that researchers and clinicians attribute resources efficiently, enabling patient access to precision treatments and ultimately improved patient outcomes.
    - The first step towards this is clear, consistent and standardized reporting of testing and mutation rates in the scientific literature.
- In addition, we invite clinicians to explore *BRAF* testing and mutation rates using the dashboard accessible from this link. This user-friendly presentation of epidemiological information may encourage the testing of patients for *BRAF* mutations, a necessary step towards optimized treatment with precision medicine.

Acknowledgements

The authors thank Brad Rosbrook and William Duggan of Pfizer for their support with the statistical analysis. The authors also thank Maxandre Jacqueline, MRes https://orcid.org/0000-0001-6465-5135 and Kim Wager, PhD https://orcid.org/0000-0001-8107-9600 of Oxford PharmaGenesis, Oxford, UK for providing medical writing support, which has been funded by Pfizer, New York, NY, USA, in accordance with Good Publication Practice 3 (GPP3) guidelines (https://www.ismpp.org/gpp3).

References

1. Brown, N.A. and K.S.J. Elenitoba-Johnson, *Enabling Precision Oncology Through Precision Diagnostics.* Annu Rev Pathol, 2020. **15**: p. 97-121.
2. Chen, H.-Z., R. Bonneville, and S. Roychowdhury, *Implementing precision cancer medicine in the genomic era.* Seminars in Cancer Biology, 2019. **55**: p. 16-27.
3. Horgan, D., et al., *Bringing Greater Accuracy to Europe's Healthcare Systems: The Unexploited Potential of Biomarker Testing in Oncology.* Biomedicine Hub, 2020. **5**(3): p. 1-42.
4. European Society of Medical Oncology, *Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Available at, https://www.esmo.org/content/download/347819/6934778/1/ESMO-CPG-mNSCLC-15SEPT2020.pdf (accessed November 2021).* 2020.
5. National Comprehensive Cancer Network, *Non-Small Cell Lung Cancer (version 7.2021). Available at, https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf (accessed November 2021 - login required).* 2021.
6. Li, M.M., et al., *Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists.* J Mol Diagn, 2017. **19**(1): p. 4-23.
7. Schork, N.J., *Randomized clinical trials and personalized medicine: A commentary on deaton and cartwright.* Soc Sci Med, 2018. **210**: p. 71-73.
8. Whiting, P.F., et al., *A systematic review classifies sources of bias and variation in diagnostic test accuracy studies.* Journal of Clinical Epidemiology, 2013. **66**(10): p. 1093-1104.

9. Blanchard, A., *Mapping ethical and social aspects of cancer biomarkers.* N Biotechnol, 2016. **33**(6): p. 763-772.
10. Dearden, S., et al., *Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap).* Annals of Oncology, 2013. **24**(9): p. 2371-2376.
11. Ha, V.T.D., J. Frizzo-Barker, and P. Chow-White, *Adopting clinical genomics: a systematic review of genomic literacy among physicians in cancer care.* BMC medical genomics, 2018. **11**(1): p. 18-18.
12. El-Galaly, T.C., C.Y. Cheah, and D. Villa, *Real world data as a key element in precision medicine for lymphoid malignancies: potentials and pitfalls.* Br J Haematol, 2019. **186**(3): p. 409-419.
13. Stenzl, A., et al., *Application of artificial intelligence to overcome clinical information overload in urological cancer.* BJU Int, 2021.
14. Guo, J., et al., *Automated Chemical Reaction Extraction from Scientific Literature.* J Chem Inf Model, 2021.
15. Burgard, T. and A. Bittermann, *Reducing Literature Screening Workload With Machine Learning.* Zeitschrift für Psychologie, 2023. **231**(1): p. 3-15.
16. Qin, X., et al., *Natural language processing was effective in assisting rapid title and abstract screening when updating systematic reviews.* Journal of Clinical Epidemiology, 2021. **133**: p. 121-129.
17. Marshall, I.J. and B.C. Wallace, *Toward systematic review automation: a practical guide to using machine learning tools in research synthesis.* Systematic Reviews, 2019. **8**(1): p. 163.
18. Kharawala, S., A. Mahajan, and P. Gandhi, *Artificial intelligence in systematic literature reviews: a case for cautious optimism.* Journal of Clinical Epidemiology, 2021. **138**: p. 243.
19. Ouyang, L., et al., *Training language models to follow instructions with human feedback.* arXiv preprint arXiv:2203.02155, 2022.
20. Brown, T., et al., *Language models are few-shot learners.* Advances in neural information processing systems, 2020. **33**: p. 1877-1901.
21. Aydın, Ö. and E. Karaarslan, *OpenAI ChatGPT generated literature review: Digital twin in healthcare.* Available at SSRN 4308687, 2022.
22. Zielinski, C., et al. *Chatbots, ChatGPT, and Scholarly Manuscripts: WAME Recommendations on ChatGPT and Chatbots in Relation to Scholarly Publications*. 2023.
23. *Tools such as ChatGPT threaten transparent science; here are our ground rules for their use.* Nature, 2023. **613**(7945): p. 612.
24. Azaria, A., *ChatGPT Usage and Limitations*. 2022.
25. Glaese, A., et al., *Improving alignment of dialogue agents via targeted human judgements.* arXiv preprint arXiv:2209.14375, 2022.
26. Lee, J., et al., *BioBERT: a pre-trained biomedical language representation model for biomedical text mining.* Bioinformatics, 2019. **36**(4): p. 1234-1240.
27. Wagner, G., R. Lukyanenko, and G. Paré, *Artificial intelligence and the conduct of literature reviews.* Journal of Information Technology, 2021: p. 02683962211048201.
28. Edwards, M.L., et al., *Real-world BRAF testing patterns for patients with metastatic colorectal cancer (mCRC), metastatic melanoma, and non-small cell lung cancer (NSCLC).* Journal of Clinical Oncology, 2021. **39**(15_suppl): p. e18778-e18778.
29. Takeda, H. and Y. Sunakawa, *Management of BRAF Gene Alterations in Metastatic Colorectal Cancer: From Current Therapeutic Strategies to Future Perspectives.* Frontiers in Oncology, 2021. **11**: p. 602194.
30. Vanni, I., et al., *The Current State of Molecular Testing in the BRAF-Mutated Melanoma Landscape.* Frontiers in Molecular Biosciences, 2020. **7**: p. 113.
31. Cheng, L., et al., *Molecular testing for BRAF mutations to inform melanoma treatment decisions: a move toward precision medicine.* Modern Pathology, 2018. **31**(1): p. 24-38.

32. Vanni, I., et al., *The Current State of Molecular Testing in the BRAF-Mutated Melanoma Landscape.* Frontiers in molecular biosciences, 2020. **7**: p. 113-113.
33. Kopetz, S., et al., *Global BRAF testing practices in metastatic colorectal cancer.* Journal of Clinical Oncology, 2021. **39**(3_suppl): p. 128-128.
34. Potter, A.J., et al., *Pathologist initiated reflex BRAF mutation testing in metastatic melanoma: experience at a specialist melanoma treatment centre.* Pathology, 2022. **54**(5): p. 526-532.
35. Boehmer, L., et al., *Identifying barriers to equitable biomarker testing in underserved patients with NSCLC: A mixed-methods study to inform quality improvement opportunities.* Journal of Clinical Oncology, 2021. **39**(28_suppl): p. 123-123.
36. Bruno, D.S., et al., *Racial disparities in biomarker testing and clinical trial enrollment in non-small cell lung cancer (NSCLC).* Journal of Clinical Oncology, 2021. **39**(15_suppl): p. 9005-9005.
37. Landry, L.G., et al., *Lack Of Diversity In Genomic Databases Is A Barrier To Translating Precision Medicine Research Into Practice.* Health Aff (Millwood), 2018. **37**(5): p. 780-785.
38. Takeda, H. and Y. Sunakawa, *Management of BRAF Gene Alterations in Metastatic Colorectal Cancer: From Current Therapeutic Strategies to Future Perspectives.* Frontiers in Oncology, 2021. **11**(942).
39. Shi, C., et al., *BRAFV600E mutation, BRAF-activated long non-coding RNA and miR-9 expression in papillary thyroid carcinoma, and their association with clinicopathological features.* World Journal of Surgical Oncology, 2020. **18**(1): p. 145.
40. Stagni, C., et al., *BRAF Gene Copy Number and Mutant Allele Frequency Correlate with Time to Progression in Metastatic Melanoma Patients Treated with MAPK Inhibitors.* Mol Cancer Ther, 2018. **17**(6): p. 1332-1340.
41. DerSimonian, R. and N. Laird, *Meta-analysis in clinical trials.* Controlled Clinical Trials, 1986. **7**(3): p. 177-188.
42. Neyeloff, J.L., S.C. Fuchs, and L.B. Moreira, *Meta-analyses and forest plots using a microsoft excel spreadsheet: step-by-step guide focusing on descriptive data analysis.* BMC Research Notes, 2012. **5**(1): p. 52.
43. Bernoulli, J., *Ars coniectandi*. 1713: Impensis Thurnisiorum, fratrum.
44. Stagni, C., et al., *BRAF Gene Copy Number and Mutant Allele Frequency Correlate with Time to Progression in Metastatic Melanoma Patients Treated with MAPK Inhibitors.* Molecular Cancer Therapeutics, 2018. **17**(6): p. 1332-1340.
45. Bode, A.M. and Z. Dong, *Recent advances in precision oncology research.* npj Precision Oncology, 2018. **2**(1): p. 11.
46. Krzyszczyk, P., et al., *The growing role of precision and personalized medicine for cancer treatment.* Technology, 2018. **6**(3-4): p. 79-100.