IMAGING

original reports

Semiautomated Pipeline to Quantify Tumor Evolution From Real-World Positron Emission Tomography/Computed Tomography Imaging

Daniel Abler, PhD[1,2]; Perrine Courlet, PhD[1,3]; Matthieu Dietz, MD[4,5]; Roberto Gatta, PhD[1]; Pascal Girard, PhD[6]; Alain Munafo, PhD[6]; Alexandre Wicky, MD[1]; Mario Jreige, MD[4]; Monia Guidi, PhD[3,7]; Sofiya Latifyan, MD[8]; Rita De Micheli, MD[8]; Chantal Csajka, PhD[3,9,10]; John O. Prior, MD, PhD[4]; Olivier Michielin, MD, PhD[1]; Nadia Terranova, PhD[6]; and Michel A. Cuendet, PhD[1,11,12]

abstract

PURPOSE A semiautomated pipeline for the collection and curation of free-text and imaging real-world data (RWD) was developed to quantify cancer treatment outcomes in large-scale retrospective real-world studies. The objectives of this article are to illustrate the challenges of RWD extraction, to demonstrate approaches for quality assurance, and to showcase the potential of RWD for precision oncology.

METHODS We collected data from patients with advanced melanoma receiving immune checkpoint inhibitors at the Lausanne University Hospital. Cohort selection relied on semantically annotated electronic health records and was validated using process mining. The selected imaging examinations were segmented using an automatic commercial software prototype. A postprocessing algorithm enabled longitudinal lesion identification across imaging time points and consensus malignancy status prediction. Resulting data quality was evaluated against expert-annotated ground-truth and clinical outcomes obtained from radiology reports.

RESULTS The cohort included 108 patients with melanoma and 465 imaging examinations (median, 3; range, 1-15 per patient). Process mining was used to assess clinical data quality and revealed the diversity of care pathways encountered in a real-world setting. Longitudinal postprocessing greatly improved the consistency of image-derived data compared with single time point segmentation results (classification precision increased from 53% to 86%). Image-derived progression-free survival resulting from postprocessing was comparable with the manually curated clinical reference (median survival of 286 *v* 336 days, $P = .89$).

CONCLUSION We presented a general pipeline for the collection and curation of text- and image-based RWD, together with specific strategies to improve reliability. We showed that the resulting disease progression measures match reference clinical assessments at the cohort level, indicating that this strategy has the potential to unlock large amounts of actionable retrospective real-world evidence from clinical records.

JCO Clin Cancer Inform 7:e2200126. © 2023 by American Society of Clinical Oncology

INTRODUCTION

The use of real-world data (RWD) to generate real-world evidence (RWE) is attracting increasing interest to complement evidence from randomized clinical trials (RCTs). Although RCTs are generally considered to provide the highest level of evidence, RWD analyses allow the investigation of a larger and more heterogeneous patient population that better reflects the diversity of diseases and treatment sequences in the clinical setting.[1-4] Increasingly personalized and individually optimized treatment approaches demand fine-grained assessment of efficacy and safety outcomes according to multiple patient traits and response patterns. Precision oncology, for example, seeks to optimize cancer therapies by understanding the patient-specific determinants of treatment response. This requires large data sets of patient characteristics, treatments, and outcome information that enable reconstructing an individual's clinical trajectory and response pattern.

Immune checkpoint inhibitors (ICIs) represent a class of immunotherapy treatments that have revolutionized the treatment of metastatic cancers, such as melanoma.[5] Despite the success of this treatment in a subgroup of the patient population, a significant proportion of patients does not respond or experiences severe side effects. Hence, a strong need remains to improve patient stratification and to better understand different patterns of response,[6-8] such as dissociate response and pseudoprogression. Several studies investigated

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on February 3, 2023 and published at ascopubs.org/journal/cci on May 5, 2023: DOI https://doi.org/10.1200/CCI.22.00126

CONTEXT

Key Objective

To demonstrate the successful implementation and identify associated challenges of a semiautomated real-world oncologic data collection and analysis pipeline for systematic quantification of tumor lesion dynamics and clinical outcomes on the basis of radiologic evidence from routine clinical practice.

Knowledge Generated

The proposed pipeline allows tapping into large amounts of retrospective real-world radiologic evidence on the evolution of metastatic cancers, otherwise only accessible through manual inspection by highly trained experts. A wide variety of quality assurance steps demonstrate the ability of such a method to provide accurate data for population-based studies.

Relevance

There is an increasing demand for real-world evidence to guide precision oncology. Rich and complex data such as clinical positron emission tomography/computed tomography images can only be unlocked at large scale with appropriate automatic tools, such as the pipeline presented here.

the efficacy and safety outcomes of ICIs in both RCTs and real-world settings. They not only partially reported similar results for both approaches[9-14] but also observed differences, mainly indicating lower survival and more severe adverse events in real-world populations.[15-18] Acknowledging the role of RWE generated from such population-based observational studies for assessing the generalizability of RCT findings, both US Food and Drug Administration and European Medicines Agency promote RWE to support regulatory decision making.[19,20]

Despite its promise and the ubiquity of clinical RWD, RWE is difficult to obtain. RWD, such as clinical or imaging data, typically must undergo labor-intensive curation steps before they can be used in quantitative population analyses and contribute to RWE. For example, RWD investigations of the response patterns after treatment with ICIs require structured access to (1) patient and treatment information for cohort selection, (2) clinical covariates such as information about blood, molecular, or genetic markers, and (3) longitudinal information about tumor burden and characteristics, ideally at the single-lesion resolution. Most of this information is not readily available from the raw data collected in clinical routine that consist of mainly unstructured entries in electronic health records (EHRs) and archived imaging data, both oriented toward clinical use. To enable the use of EHR data for RWD studies, information about patient characteristics, diagnoses, and treatment history must be extracted and structured. Similarly, imaging data need to be annotated to identify and extract relevant information. In the context of oncologic imaging for assessing the response to anticancer treatments, quantitative information about the longitudinal evolution of tumor characteristics and burden is needed. This type of information is typically collected in RCTs using criteria such as RECIST[21] or positron emission tomography (PET) Response Criteria in Solid Tumors.[22] Both rely on longitudinal measurements of tumor size or quantification of tumor metabolism from imaging modalities such as computed tomography (CT), PET, or a combination of both (PET/CT). Although lesion-level quantitative PET/CT assessment is also performed in clinical practice, criteria such as RECIST are not directly used to assess patient response or disease progression and the patient's overall tumor burden remains unquantified. Instead, radiologic evidence is mostly described qualitatively in natural language, thus providing insufficient details for subsequent quantitative assessment of tumor response.[23]

Furthermore, the validity of RWE strongly relies on data quality.[1] As manual curation is very time-consuming, large-scale population analyses require automated RWD processing approaches. For such strategies to produce actionable RWE, automation-induced inaccuracies must be minimized, and any resulting errors need to be well-characterized and considered when interpreting results. If performed carefully, automated curation of textual clinical data from EHRs can yield results comparable with manually curated data, as recently demonstrated for textual data from EHR in the context of population pharmacokinetics.[24]

This study provides further evidence for the feasibility of automated RWD curation in the context of response assessment to ICIs. Beginning by outlining the conceptual steps involved in RWD extraction and curation for RWE creation, we propose strategies for quality control and iterative refinement of automatically curated text- and image-based RWD. In particular, our work focuses on strategies for assessing the suitability of automated longitudinal tumor burden measurements for population-based analyses of tumor response to ICIs. We demonstrate how longitudinal information can be exploited to increase the accuracy of automated tumor burden assessments. We further show that the resulting information can serve as a substitute for manually curated disease progression assessments at the population level.

METHODS

Figure 1 illustrates the key steps involved in semiautomated collection and curation of clinical and imaging RWD for

retrospective biomedical studies. These include (1) extraction of RWD from hospital sources, (2) enrichment of the raw data with information that enable identification and retrieval of key data elements, (3) data selection on the basis of this information, and (4) quality assurance to assess data consistency. In practice, steps (2)-(4) form an iterative data consolidation process that terminates when the targeted consistency criteria have been met. Despite differences in the specific methods and strategies, the same conceptual sequence of steps applies to different types of RWD, such as text-based information from EHRs or clinical imaging data. Although quality assurance steps are typically supported by methodological and technical developments tailored to specific RWD use cases (local clinical practice, information system, data formats, etc), we focus here on generalizable aspects in the context of text- and image-based RWD curation for oncologic research.

Text Mining and Manual Annotation

Extraction of text-based RWD relied on a data aggregation and semantic annotation system previously developed at the Lausanne University Hospital (CHUV) Precision Oncology Center and enriched during this project. This tool helps data managers capture and structure clinical data within a graphical user interface that integrates text mining and natural language processing algorithms. The structured data (eg, treatment line or radiologic response) are then quality-controlled by clinicians and complemented with structured EHR data such as laboratory values and radiotherapy administration data.

Cohort and imaging examination selection. The present RWD exploration focuses on the first ICI line in patients with metastatic cutaneous melanoma treated at CHUV between April 2014 and February 2021. Usage of patient data was approved by the local Ethics Committee (CER-VD protocol No. 2019-00448) for patients who did not refuse consent. For patients to be included in this cohort, at least two PET/CT images between three months before ICI initiation and one year after ICI discontinuation or the next treatment line were required. In the following, we refer to images acquired before and after first-line ICI start as baseline and follow-up images, respectively. To prevent confounders from obscuring real ICI effect evaluation, localized treatments such as surgery or radiation therapy were considered when selecting the appropriate imaging examinations. We excluded baseline images collected before and follow-up images collected after cancer-related surgeries. Similarly, images collected after the first radiotherapy session (excluding treatment for brain metastases since not evaluable using PET/CT) were omitted. For the selected cohort, demographic (eg, age, sex) and clinical characteristics (eg, Eastern Cooperative Oncology Group performance status), laboratory values, and somatic mutations (ie, *BRAF*, *NRAS*) were collected.

Quality assurance. We verified and corrected the clinical data set using multiple quality control mechanisms and by involving clinicians to help validate the extracted data. In particular, we used process mining, an important novel methodology for visualization and deep analysis of event-based data sets, as implemented in the pMineR R library.[25]

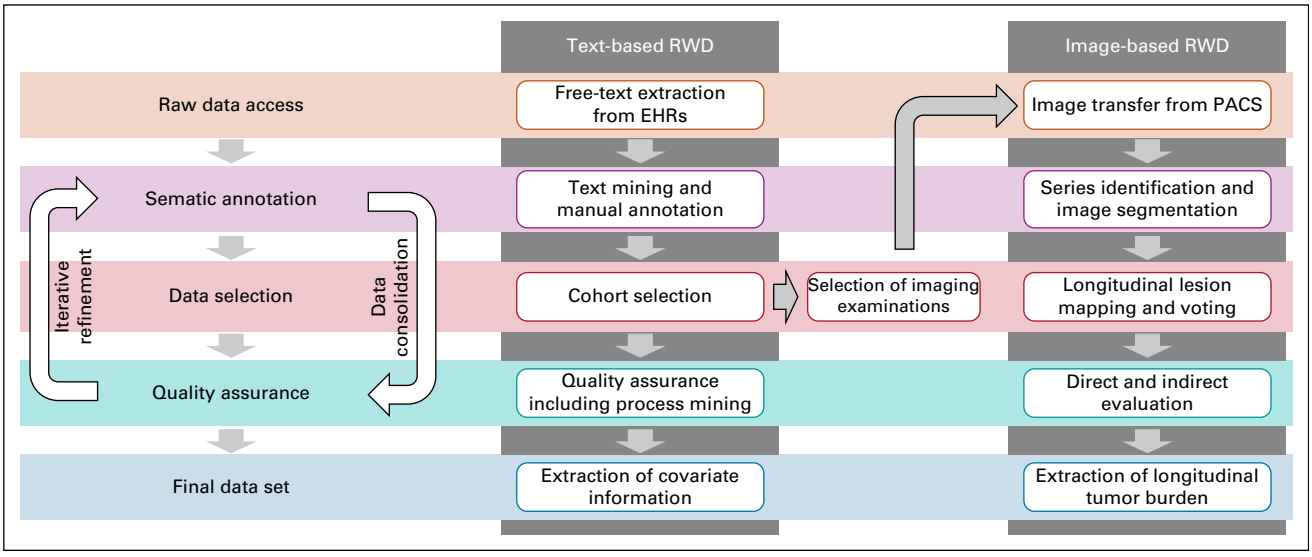


FIG 1. Key steps involved in collection and curation of real-world clinical and imaging data. Semantic annotation of raw data enables querying for key cohort selection criteria. Quality assurance assesses the internal consistency of the selected data set to detect and correct shortcomings of the automated annotation process. The iterative data consolidation process of semantic annotation, data selection, and quality assurance terminates when the targeted consistency criteria have been met. The same conceptual sequence of steps applies to the curation of different types of RWD, such as text-based information from EHRs or clinical imaging data. EHR, electronic health record; PACS, Picture Archiving and Communication System; RWD, real-world data.

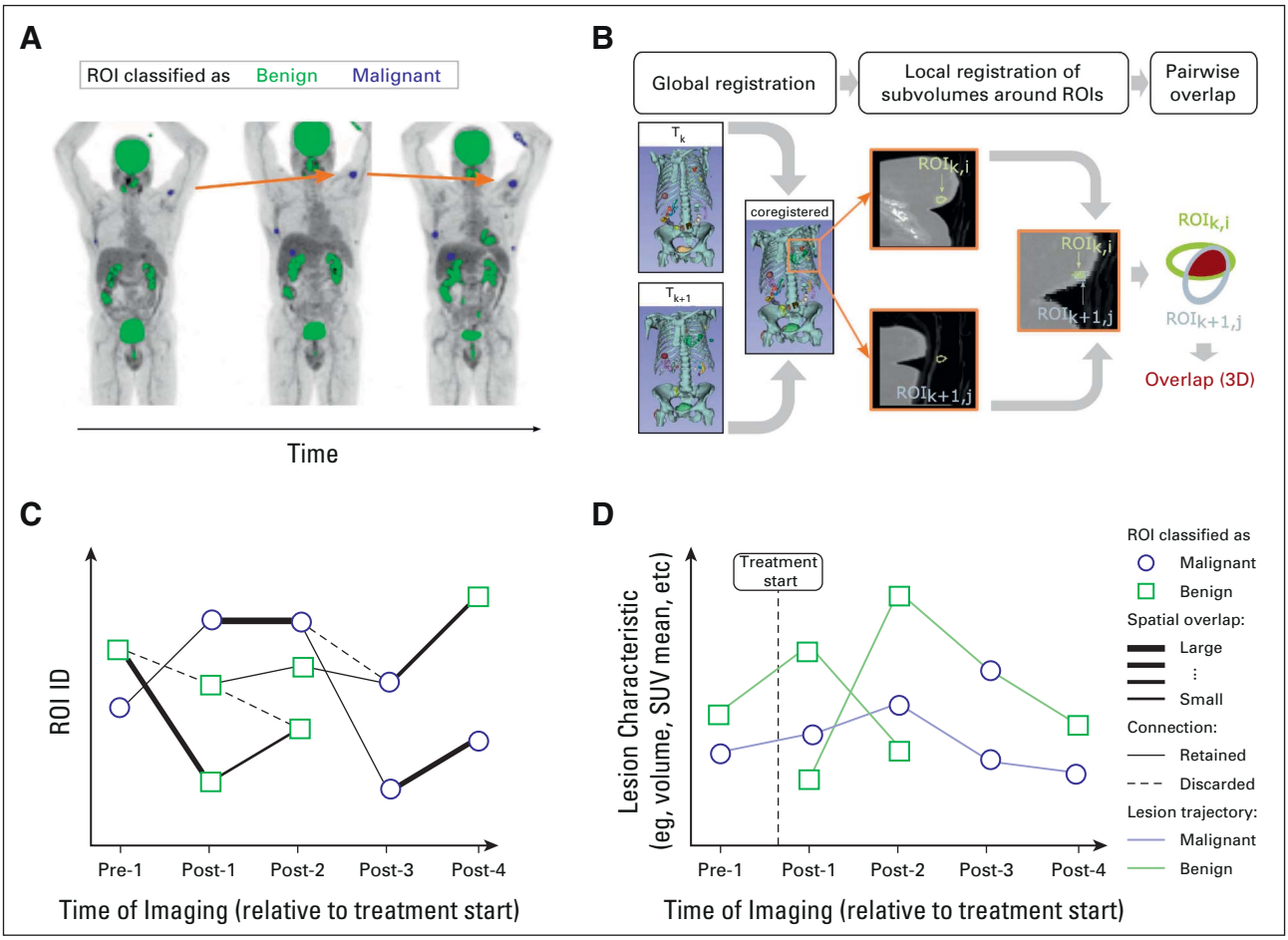


FIG 2. Longitudinal follow-up of individual tumor lesions: (A) longitudinal mapping aims at establishing correspondences between the ROIs observed and segmented across a patient's imaging examinations. (B) ROIs observed at successive imaging time points are aligned by global and local registration steps. Spatial overlap in a shared frame of reference indicates the likelihood of two ROIs from successive time points to correspond to the same tumor lesion. (C) The spatiotemporal relationship between all ROIs observed in a patient's imaging examinations are captured in a graph-based data structure. Retaining/discarding graph edges on the basis of their weights (magnitude of overlap) enables identification of longitudinally corresponding ROIs. (D) Evolution of lesion characteristics (eg, total metabolic tumor volume) and classification status (malignant/benign) of longitudinally mapped ROIs. Inconsistent classification of longitudinally mapped ROIs of the same tumor lesion are resolved by majority voting, that is, assigned according to the most common (>50%) classification status across all the lesion's observations. 3D, three-dimensional; ROI, region of interest; SUV, standardized uptake values.

For this, an event log was generated, listing the sequence of events for each patient. Then, a first-order Markov process model was inferred, which allows quick inconsistency detection by spotting aberrant paths in the process graphs.[26]

Curation of Imaging Data

Automated series identification and image segmentation. To select the PET and CT images best suited for further processing, we established rules that automatically identified the relevant acquisition protocols among all images retrieved from the institutional Picture Archiving and Communication System (PACS). Segmentation of the selected PET/CT images relied on the artificial intelligence–based *PET-Assisted Reporting System* (PARS) prototype software (Siemens Medical Solutions USA Inc, Knoxville, TN).[27,28] PARS segments and identifies tumor lesions from PET/CT in a two-step process: first, three-dimensional regions of interest (ROIs) with elevated Standardized Uptake Values (SUVs) are selected. These are then classified into (1) *benign* ROIs where elevated SUV is physiologically expected (eg, metabolically active regions such as the brain) and (2) *malignant* ROIs where high SUV indicates the presence of metabolically active cancer cells. In addition, each ROI is assigned to an anatomic location, selected from a set of predefined body parts and organs. The assignment of malignancy status and anatomic location relies on convolutional neural networks trained on lymphoma and lung cancer cases. For these indications, PARS is reported to achieve an AUC of the receiver-operating characteristic of >0.95 for lesion classification and anatomic localization, and accuracies of 96% for body part and 87% for organ localization.[27]

Longitudinal lesion mapping. Additional processing is needed to identify lesion correspondences across imaging time points (Fig 2A). We developed an automated approach for this longitudinal mapping task on the basis of the assumption that ROIs repeatedly observed at the same location likely correspond to the *same* tumor lesion. The mapping started with a sequence of image registration steps (Fig 2B). First, images were cropped to the patient's trunk, and the CT images corresponding to two subsequent acquisition time points were spatially aligned by affine registration. Then, the subvolumes surrounding each ROI at time points k and k + 1 were extracted, and deformable registration was performed between all pairs ($ROI_{i,k}$, $ROI_{j,k+1}$). For each coregistered pair, we computed the spatial overlap between ROIs. From the resulting information, a graph data structure was built (Fig 2C), which enabled identification of corresponding ROIs across imaging examinations and thus longitudinal follow-up of individual tumor lesions (Fig 2D).

As longitudinal mapping was performed independent of ROI malignancy status, the resulting longitudinal correspondences may contain ROIs of benign and malignant classification outcome (Fig 2D), a potential inconsistency resulting from the uncertainty involved in the lesion classification process. However, we could leverage this information to improve the malignancy status prediction for the entire lesion trajectory by performing majority voting on the basis of the malignancy status of its individual observations.

Lesion-level and patient-level performance evaluation. Visual inspection of PET/CT segmentations revealed the presence of misdetected or misclassified lesions, as previously reported.[29] For example, catheters for ^{18}F-labeled fluorodeoxyglucose injection were frequently classified as *malignant* skin lesions by PARS because of their elevated SUV and placement in proximity to the skin (Data Supplement). To evaluate automatic segmentation and classification performance, as well as the impact of longitudinal mapping and voting on overall data quality, we established two benchmarks.

Lesion-level evaluation against expert-annotated ground truth. Two independent reviews were performed by two experienced nuclear medicine experts. First, one expert reviewed PARS segmentation results of baseline and first follow-up images for a group of 10 patients, balanced in terms of patient clinical progression status (extracted from clinical records). This data set served to provide an initial estimate of segmentation quality and to adjust segmentation parameters. A second review by the other expert aimed at creating a representative ground-truth data set to allow systematic evaluation of segmentation performance and of the effect of various postprocessing approaches. For this purpose, the automatically segmented tumor lesions (n = 214) in about 5% of all imaging examinations (n = 24) were reviewed to identify ROIs erroneously classified as malignant. The examinations were randomly selected but chosen to reflect the repartition of disease burden in the overall population (quantified by the number of malignant lesions, n_{mal}): 50% with $n_{mal} \leq 5$, 25% with $5 < n_{mal} \leq 15$, and 25% with $n_{mal} > 15$. We also investigated all imaging examinations (n = 26) in which PARS had failed to detect the presence of malignant lesions despite the radiology report suggesting otherwise. On the basis of these expert reviews, we evaluated (1) classification *precision* at the lesion level and (2) the *negative predictive value* (NPV) at the examination level.

Patient-level evaluation against clinical outcome. To validate our lesion selection approach at the patient level, we evaluated the agreement between the temporal evolution of image-derived total metabolic tumor volume (TMTV) and clinical response. As clinical reference for patient response to immunotherapy, we used an in-house score that dichotomizes the information from radiologic reports into progressive disease (PD) or nonprogressive disease (NPD). We compared the clinical response with results obtained from automated image analysis in terms of TMTV change, progression events derived from TMTV change, and resulting progression-free survival (PFS). Differences in PFS were assessed using the log-rank test.

RESULTS

The cohort selection steps described earlier resulted in a cohort of 108 patients with melanoma (59% male) with a median age of 68 years at baseline. About half of the patients (n = 51) received ipilimumab and nivolumab concomitantly, potentially followed by nivolumab maintenance (n = 22). Baseline (median, 1; range, 1-2) and follow-up (median, 3; range, 1-14) images were collected for each patient. Further cohort characteristics are provided in the Data Supplement. Of 509 eligible PET/CT imaging examinations, 446 examinations were successfully retrieved from the institutional PACS, correctly identified, and segmented for organs and lesions.

Process Mining Confirms Clinical Data Quality and Reveals Patient Pathways

Figure 3 illustrates process mining on a subgroup of patients receiving pembrolizumab (n = 32). Depicting transitions between various clinical events, the figure exemplifies the diversity of care pathways encountered in a real-world setting. Process mining enabled iterative correction of data curation errors until self-consistency was achieved and provided descriptive statistics, as detailed in Figure 3.

Longitudinal Mapping and Voting Increase Precision of Malignancy Identification

Compared with expert-classified ground-truth data, we observed a classification precision of 53% and a NPV of 80% for PARS predictions, meaning that (1) only half of the lesions identified as malignant by PARS were confirmed to be malignant on expert evaluation and (2) about 80% of all examinations in which PARS did not identify any malignant ROI were confirmed to be free of tumor lesions. The low precision

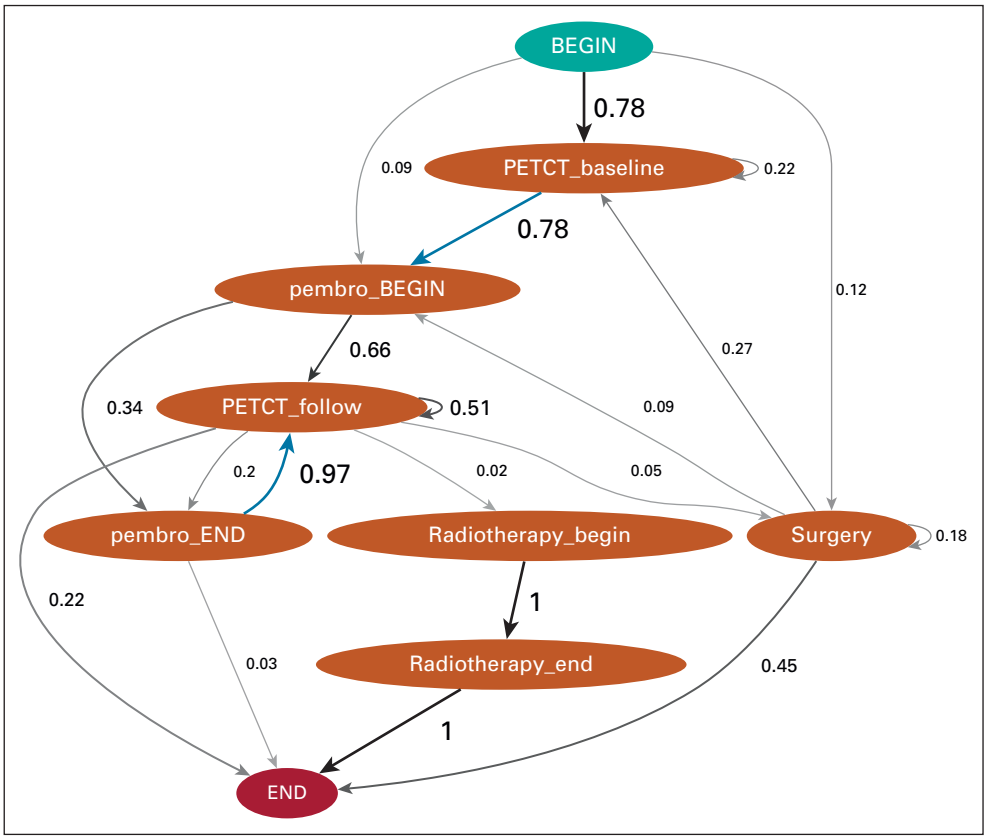


FIG 3. Process-mining example: directed graph showing the proportion of outgoing transitions between states generated using pMineR[25] for a subset of 32 patients with melanoma receiving pembrolizumab. Arrows referring to the summary statistics described below are given in blue. No direct paths from baseline imaging to surgery or from surgery to follow-up imaging were observed after data consolidation, confirming the adequate handling of surgeries during cohort preprocessing. In addition to quality control, process mining provides descriptive statistics, showing that treatment is initiated directly after one baseline PET/CT in 78% of cases. This relatively low proportion can be attributed to the availability of other image modalities (ie, CT or magnetic resonance images) for cancer diagnosis and to imaging performed outside of CHUV. In our time interval of interest, all patients had at least one follow-up PET/CT after the last pembrolizumab administration. NB: The denominator for the calculation of the proportions is the number of patients in the upstream node, not the total number of patients. CT, computed tomography; pembro, pembrolizumab; pembro_BEGIN, pembrolizumab initiation; pembro_END, last pembrolizumab administration in the considered time interval (does not necessarily correspond to the actual end of treatment, especially because of censored data after relevant surgeries and radiotherapy sessions); PET, positron emission tomography.

is further evidenced in the longitudinal evolution of tumor burden (Fig 4A) where it results in erratic lesion trajectories.

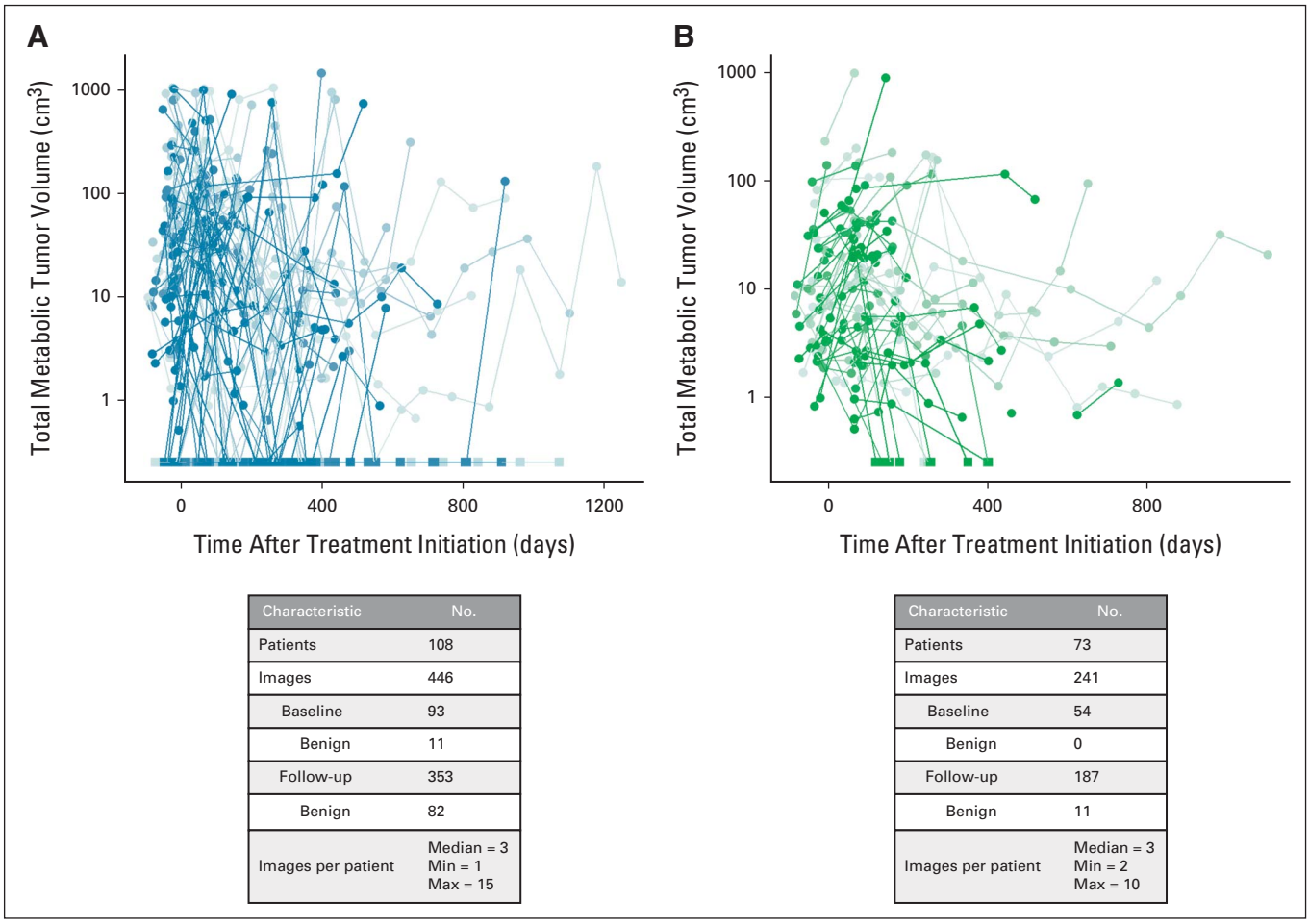
A higher precision of 66% was achieved for the subset of tumor lesions retained after application of longitudinal mapping. Indeed, improved precision in this subset of lesions (n = 85) is expected because longitudinal mapping discards lesions that were observed at a single time point only, which frequently correspond to incorrectly classified regions of benign high SUV uptake.

More importantly, access to individual lesion trajectories enabled a majority voting rule to be applied for determining the most likely overall status of a lesion trajectory. By reclassifying each involved observation accordingly, the overall classification precision was further increased to 86%. Figure 5 illustrates these findings and details the confusion matrix for this binary classification problem in terms of ROI counts and volume. The Data Supplement summarizes these findings in numerical form.

Combined with manual identification of truly benign baseline images, this approach greatly improved the consistency of our data set by removing outliers caused by sporadic misclassification (Fig 4B). This plot also emphasized the large interindividual variability in tumor dynamics, which is expected in patients receiving ICIs.[30,31]

Image-Derived Tumor Dynamics Is Consistent With Clinical Response Assessment

Figure 6A compares the relative change in TMTV derived from imaging with the clinical progression status



Characteristic	No.
Patients	108
Images	446
Baseline	93
Benign	11
Follow-up	353
Benign	82
Images per patient	Median = 3 Min = 1 Max = 15

Characteristic	No.
Patients	73
Images	241
Baseline	54
Benign	0
Follow-up	187
Benign	11
Images per patient	Median = 3 Min = 2 Max = 10

FIG 4. Tumor dynamics represented as TMTV as a function of time after treatment initiation, associated with summary statistics. (A) Tumor dynamics on the basis of PARS raw segmentation results. (B) Tumor dynamics after data consolidation using longitudinal mapping, majority voting, and manual identification of truly benign baseline images (identified by comparison of baseline images without detected tumor burden with radiologic reports from clinical care). Observations for the same patient are joined by a solid line and patients are represented by different color shades. TMTV of images without detected tumor burden was imputed to 0.25 cm^3 (half of the lower limit of quantification of PARS) and is represented by squares. max, maximum; min, minimum; PARS, *PET-Assisted Reporting System*; TMTV, total metabolic tumor volume.

(PD, NPD) for these examinations. It distinguishes the change in TMTV on the basis of raw segmentation results and longitudinal mapping and voting. Although the observed median volume change is consistent with the clinical progression status in most cases and only shows little dependence on the processing approach, longitudinal mapping reduces the variability of these distributions. Longitudinal voting also resulted in improved agreement between image-derived and clinical progression events with a median agreement of 85% compared with a median agreement of 62% without this postprocessing step (Fig 6B). Similarly, image-derived progression events on the basis of raw and postprocessed segmentation results yielded different median PFS times with a median of 175 and 286 days, respectively ($P < .01$). Postprocessed segmentation results closely reproduced the clinically derived PFS (median of 336 days, $P = .89$), whereas raw segmentation results did not ($P < .01$; Fig 6C). These findings are summarized in the Data Supplement.

DISCUSSION

Efficient data curation and quality control are central challenges in RWD research. No standard approaches for automatic RWD extraction and curation exist, and hence, RWD curation is performed in a costly manual process to increase data quality for RWD studies. Here, we argue for the need for improved systems that can leverage RWD and we present evidence for the feasibility of automated RWD curation in the context of tumor response evaluation to ICIs. We identified recurring steps in the collection and curation process of text- and image-based RWD and proposed a general framework and specific strategies for addressing them. We showed that process mining can be integrated in iterative data consolidation workflows to improve data quality by providing easily interpretable synthetic

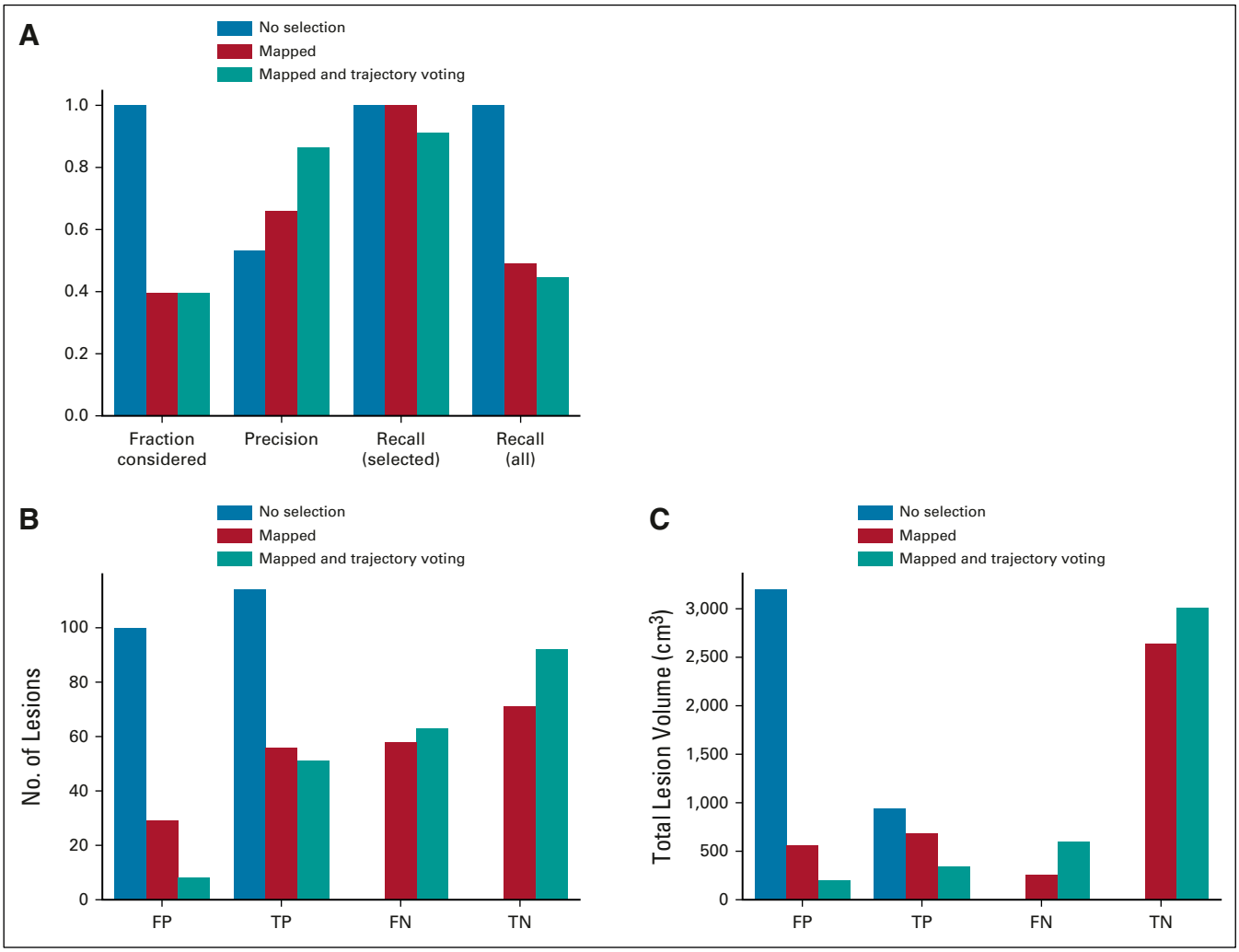


FIG 5. Lesion-level evaluation of automatic image analysis results against expert-annotated subset of lesion segmentations (n = 214). (A) Mapping and longitudinal voting increase precision, without substantially affecting recall within the selected subset. However, mapping reduces the fraction of lesions that can be evaluated, thus negatively affecting overall recall when nonevaluable lesions are counted as either true negatives or false negatives. (B) Number of lesions identified as FP, TP, FN, or TN explaining the trends for overall recall shown in (A). Note that mapping not only substantially increases the detection of TN (n = 71) but also results in an increased number of FN (n = 58) because of the removal of lesions by the mapping process. (C) Total volume of lesions identified as FP, TP, FN, or TN. Note that despite (B) a comparable number of FN and TN lesions, the volume of lesions correctly discarded by mapping (mapped, TN) exceeds the volume of incorrectly discarded (mapped, FN) lesions by a factor of >10. FN, false negatives; FP, false positives; TN, true negatives; TP, true positives.

representations of patient care pathways. Furthermore, we demonstrated that suitable postprocessing enables off-the-shelf PET/CT segmentation tools to quantify tumor burden from real-world imaging data, in such a way that resulting population-level outcome measures such as PFS match those derived from time-consuming expert annotations.

This study highlights the role of data selection and the importance of quality control metrics in automated RWD curation: although RWD are intrinsically limited to the information routinely documented as part of clinical care, automated collection and curation strategies may increase the risk of missing or misinterpreting existing information in a subset of patients for whom the extraction or semantic annotation algorithms are not optimally adapted. Those situations confront researchers with a trade-off between data quality and final data set size. In the presented case, we used an existing tool (PARS) for automatic tumor segmentation from PET/CT imaging. Similar to other products, PARS is designed to assist clinicians in identifying and contouring tumor lesions and prioritizes high detection sensitivity to avoid potentially malignant lesions being overlooked. In the context of quantitative response assessment, however, the software's high detection sensitivity, resulting in a high number of incorrectly detected tumors, introduced substantial noise in the data curation process. We mitigated the impact of segmentation uncertainty by applying data selection and consistency-improving schemes, thus increasing data quality while reducing the size of our data set. The

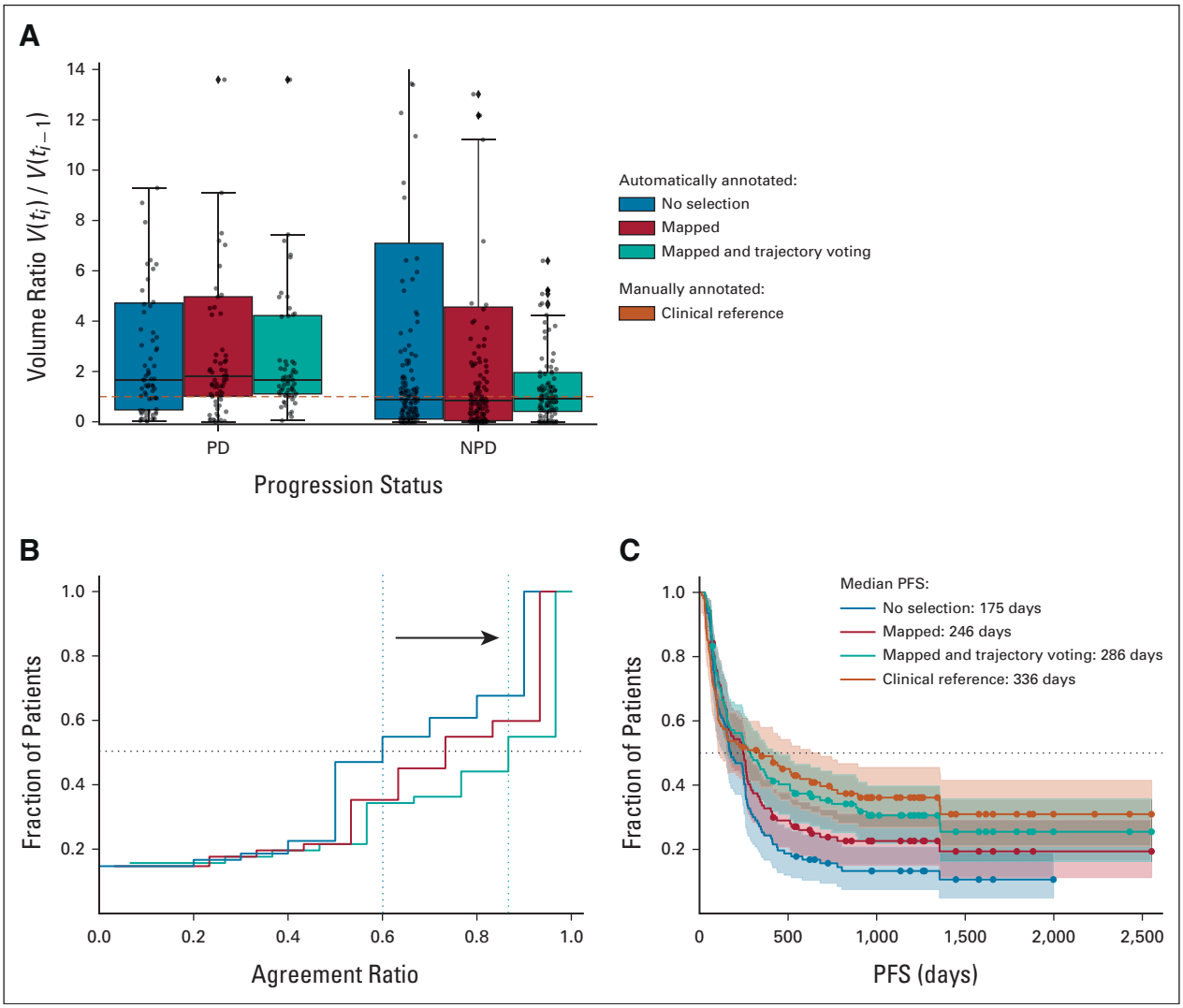


FIG 6. Evaluation of automatic image analysis results against clinically curated disease progression information: (A) ratio of total metabolic tumor volume observed at consecutive imaging examinations grouped by clinical progression status for different levels of image processing. PD is associated with a volume ratio of >1 and NPD to a ratio of ≤1. Longitudinal mapping and trajectory voting reduce the spread of observed volume ratios in the respective progression category. (B) Cumulative distribution of the agreement ratio between image-derived and clinical progression measures for each patient. Longitudinal mapping and trajectory voting result in an increase of median agreement from 62% to 85%. (C) PFS on the basis of image-derived compared with clinical progression events. Progression events on the basis of longitudinal mapping and trajectory voting resulted in extended PFS compared with events on the basis of unprocessed segmentation results (median PFS of 286 days *v* 175 days, $P < .01$). Postprocessed segmentation results thus closely reproduced the clinically curated PFS (median PFS of 336 days, $P = .89$), whereas unprocessed segmentation results did not ($P < .01$). NPD, nonprogressive disease; PD, progressive disease; PFS, progression-free survival.

proposed postprocessing strategy on the basis of longitudinal tracking of individual lesions enabled us to (1) identify the ROIs that persist over time and (2) select those that were consistently classified as malignant. Together, these criteria allowed the high number of false positives to be reduced significantly while preserving the overall treatment response pattern. We monitored the impact of curation steps on data quality and quantity to guide the iterative RWD consolidation process using three different types of assessments: (1) cohort-level self-consistency with process mining; (2) lesion-level performance evaluation against a small, but representative expert-annotated ground-truth data set; and (3) patient-level comparison of image-derived measurement with clinical progression assessments. Our evaluation indicates that the data resulting from automated RWD curation can support population-based analyses, such as PFS estimation and longitudinal response modeling.

While our study was based on 446 imaging examinations, lesion-level assessment (2) could only be performed on a subset of about 5% of these examinations. Although limited in size, this subset was chosen to be representative in terms of the number of identified lesions for patients with low, medium,

and high tumor burden, respectively. Moreover, the fact that the proposed data preparation strategy successfully reproduced clinical progression events in assessment (3) gives us strong confidence in the generalizability of our findings. Nonetheless, our pipeline should be validated in large-scale future studies including a larger number of expert-reviewed examinations of multiple cancer types, originating from multiple centers and acquired on multiple imaging platforms.

We expect the combination of longitudinal lesion tracking and consistency voting to have high potential for automated RWD curation for population-based analyses of cancer treatment outcomes, but the accuracy achievable with current automatic segmentation tools may not suffice for usage in clinical applications. Nevertheless, we believe that automatic longitudinal tracking, in conjunction with improved segmentation tools, will also have applications beyond the population research context, for example, to assist radiologists in follow-up examination assessments or as a monitoring tool that helps oncologists review patient cases on the basis of trends in disease evolution.

Although further development is needed, we believe that semi- or fully automated approaches such as the ones presented will ultimately help make RWD more accessible. By combining clinical and imaging data, these approaches pave the way for improved patient selection and treatment recommendations in the era of personalized medicine. A recent paper highlighted the potential of artificial intelligence techniques combined with pharmacometrics to guide treatment individualization in the specific context of ICIs.[32] Indeed, our curated tumor evolution data can be analyzed using pharmacokinetic/pharmacodynamic modeling and simulation approaches to capture patterns of tumor responses to treatment.[33,34] Furthermore, there is growing interest in medical imaging data to guide therapies via noninvasive image-based biomarkers. Following the segmentation and longitudinal mapping steps, quantitative image descriptors, so-called radiomics features, can be computed straightforwardly for all identified tumors. Such features were shown to be predictive markers of overall survival in several cancer types[35,36] and have the potential to predict early response to ICIs in patients with melanoma.[37] Automatic extraction and curation of imaging RWD may thus provide important opportunities, not only for the confirmation of RCT results in a real-world setting but also for biomarker discovery. The proposed framework can help collect and analyze data in a harmonized way across different hospitals to assemble larger cohorts for RWD multicentric studies.

AFFILIATIONS

[1]Department of Oncology, Precision Oncology Center, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[2]Institute of Informatics, School of Management, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
[3]Centre for Research and Innovation in Clinical Pharmaceutical Sciences, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[4]Nuclear Medicine and Molecular Imaging Department, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[5]INSERM U1060, CarMeN Laboratory, University of Lyon, Lyon, France
[6]Translational Medicine, Merck Institute of Pharmacometrics, Lausanne, Switzerland, an Affiliate of Merck KGaA, Darmstadt, Germany
[7]Service of Clinical Pharmacology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[8]Service of Medical Oncology, Department of Oncology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[9]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, University of Lausanne, Geneva, Switzerland
[10]School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland
[11]Swiss Institute of Bioinformatics, University of Lausanne, Lausanne, Switzerland
[12]Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY

CORRESPONDING AUTHOR

Michel A. Cuendet, PhD, Department of Oncology, Precision Oncology Center, Lausanne University Hospital, Rue du Bugnon 25A, 1011 Lausanne, Switzerland; e-mail: michel.cuendet@chuv.ch.

EQUAL CONTRIBUTION

D.A. and P.C. contributed equally as cofirst authors to this work. N.T. and M.A.C. contributed equally as cosenior authors to this work.

SUPPORT

Supported by the Real-World Tumor Dynamics collaboration grant of the Merck KGaA (Frankfurter Straße 250, 64293 Darmstadt, Germany) and the Lausanne University Hospital (Rue du Bugnon 46, 1011 Lausanne, Switzerland).

AUTHOR CONTRIBUTIONS

Conception and design: Daniel Abler, Perrine Courlet, Roberto Gatta, Pascal Girard, Alain Munafo, Chantal Csajka, John O. Prior, Olivier Michielin, Nadia Terranova, Michel A. Cuendet
Administrative support: John O. Prior, Michel A. Cuendet
Provision of study materials or patients: Sofiya Latifyan, John O. Prior
Collection and assembly of data: Daniel Abler, Perrine Courlet, Matthieu Dietz, Alexandre Wicky, Mario Jreige, Rita De Micheli, Olivier Michielin
Data analysis and interpretation: Daniel Abler, Perrine Courlet, Monia Guidi, Sofiya Latifyan, John O. Prior, Olivier Michielin, Nadia Terranova, Michel A. Cuendet
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's

conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.
Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Daniel Abler
Research Funding: Merck KGaA (Inst)

Perrine Courlet
Employment: Merck KGaA, Debiopharm Group
Research Funding: Merck KGaA (Inst)

Roberto Gatta
Research Funding: Merck Serono (Inst)

Pascal Girard
Employment: Merck Serono

Alain Munafo
Employment: Merck KGaA
Patents, Royalties, Other Intellectual Property: A.M. is a coinventor of a patent on the way to administer cladribine orally in Multiple Sclerosis
Travel, Accommodations, Expenses: Merck KGaA

Olivier Michielin
Consulting or Advisory Role: Bristol Myers Squibb, MSD, Novartis, Roche, Amgen, Pierre Fabre
Research Funding: MSD, Bristol Myers Squibb, NeraCare GmbH
Expert Testimony: Bristol Myers Squibb
Travel, Accommodations, Expenses: Bristol Myers Squibb, MSD

Nadia Terranova
Employment: Merck Serono S.A
Stock and Other Ownership Interests: Merck Serono

Michel A. Cuendet
Research Funding: Merck Serono (Inst), Bristol Myers Squibb/Pfizer (Inst)

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

We thank the Clinical Research Data Warehouse team of the CHUV for providing access to the clinical data, Marian Caikovski (CHUV) for developing our data curation tools, and Frédéric Pédron (CHUV) for assisting with PET/CT image extraction. We are grateful to Bénédicte Maréchal, Fabian Anderegg, Bruce Spottiswoode, and Vijay Shah (Siemens) for granting access to and providing support for the PARS software prototype.

REFERENCES

1. Booth CM, Karim S, Mackillop WJ: Real-world data: Towards achieving the achievable in cancer care. Nat Rev Clin Oncol 16:312-325, 2019
2. Suvarna V: Real world evidence (RWE)—Are we (RWE) ready? Perspect Clin Res 9:61-63, 2018
3. Makady A, de Boer A, Hillege H, et al: What is real-world data? A review of definitions based on literature and stakeholder interviews. Value Health 20:858-865, 2017
4. Sherman RE, Anderson SA, Dal Pan GJ, et al: Real-world evidence—What is it and what can it tell us? N Engl J Med 375:2293-2297, 2016
5. Carlino MS, Larkin J, Long GV: Immune checkpoint inhibitors in melanoma. Lancet 398:1002-1014, 2021
6. Humbert O, Chardin D: Dissociated response in metastatic cancer: An atypical pattern brought into the spotlight with immunotherapy. Front Oncol 10:566297, 2020
7. Borcoman E, Kanjanapan Y, Champiat S, et al: Novel patterns of response under immunotherapy. Ann Oncol 30:385-396, 2019
8. Jia W, Gao Q, Han A, et al: The potential mechanism, recognition and clinical significance of tumor pseudoprogression after immunotherapy. Cancer Biol Med 16:655-670, 2019
9. Areses Manrique MC, Mosquera Martínez J, García González J, et al: Real world data of nivolumab for previously treated non-small cell lung cancer patients: A Galician lung cancer group clinical experience. Transl Lung Cancer Res 7:404-415, 2018
10. Ascierto PA, Simeone E, Sileni V, et al: Clinical experience with ipilimumab 3 mg/kg: Real-world efficacy and safety data from an expanded access programme cohort. J Transl Med 12:116, 2014
11. Asher N, Ben-Betzalel G, Lev-Ari S, et al: Real world outcomes of ipilimumab and nivolumab in patients with metastatic melanoma. Cancers (Basel) 12:2329, 2020
12. Lin S-Y, Yang C-Y, Liao B-C, et al: Tumor PD-L1 expression and clinical outcomes in advanced-stage non-small cell lung cancer patients treated with nivolumab or pembrolizumab: Real-world data in taiwan. J Cancer 9:1813-1820, 2018
13. Noronha V, Abraham G, Patil V, et al: A real-world data of Immune checkpoint inhibitors in solid tumors from India. Cancer Med 10:1525-1534, 2021
14. Pasello G, Pavan A, Attili I, et al: Real world data in the era of immune checkpoint inhibitors (ICIs): Increasing evidence and future applications in lung cancer. Cancer Treat Rev 87:102031, 2020
15. Ghisoni E, Wicky A, Bouchaab H, et al: Late-onset and long-lasting immune-related adverse events from immune checkpoint-inhibitors: An overlooked aspect in immunotherapy. Eur J Cancer 149:153-164, 2021
16. Khozin S, Carson KR, Zhi J, et al: Real-world outcomes of patients with metastatic non-small cell lung cancer treated with programmed cell death protein 1 inhibitors in the year following U.S. regulatory approval. Oncologist 24:648-656, 2019
17. La J, Cheng D, Brophy MT, et al: Real-world outcomes for patients treated with immune checkpoint inhibitors in the veterans affairs system. JCO Clin Cancer Inform 4:918-928, 2020
18. de Ávila Machado MA, de Moura CS, Chan K, et al: Real-world analyses of therapy discontinuation of checkpoint inhibitors in metastatic melanoma patients. Sci Rep 10:14607, 2020
19. Observational Data (Real World Data). London, UK, European Medicines Agency, 2017
20. Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics. Silver Spring, MD, US Food and Drug Administration, 2019
21. Eisenhauer EA, Therasse P, Bogaerts J, et al: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer 45:228-247, 2009
22. Wahl RL, Jacene H, Kasamon Y, et al: From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors. J Nucl Med 50:122S-150S, 2009 (suppl 1)
23. Griffith SD, Tucker M, Bowser B, et al: Generating real-world tumor burden endpoints from electronic health record data: Comparison of RECIST, radiology-anchored, and clinician-anchored approaches for abstracting real-world progression in non-small cell lung cancer. Adv Ther 36:2122-2136, 2019
24. Choi L, Beck C, McNeer E, et al: Development of a system for postmarketing population pharmacokinetic and pharmacodynamic studies using real-world data from electronic health records. Clin Pharmacol Ther 107:934-943, 2020

25. Gatta R, Lenkowicz J, Vallati M, et al: pMineR: An innovative R library for performing process mining in medicine, in ten Teije A et al (eds.): Artificial Intelligence in Medicine: Lecture Notes in Computer Science. Cham, Switzerland, Springer International, 2017, pp 351-55
26. Gatta R, Vallati M, Lenkowicz J, et al: Generating and comparing knowledge graphs of medical processes using pMineR, in Proceedings of the Knowledge Capture Conference. Austin, TX, ACM, 2017, pp 1-4
27. Sibille L, Seifert R, Avramovic N, et al: ^{18}F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks. Radiology 294:445-452, 2020
28. Capobianco N, Meignan M, Cottereau A-S, et al: Deep-learning ^{18}F-FDG uptake classification enables total metabolic tumor volume estimation in diffuse large B-cell lymphoma. J Nucl Med 62:30-36, 2021
29. Pinochet P, Eude F, Becker S, et al: Evaluation of an automatic classification algorithm using convolutional neural networks in oncological positron emission tomography. Front Med 8:628179, 2021
30. Chatterjee M, Elassaiss-Schaap J, Lindauer A, et al: Population pharmacokinetic/pharmacodynamic modeling of tumor size dynamics in pembrolizumab-treated advanced melanoma: Tumor modeling in pembrolizumab-treated melanoma. CPT Pharmacometrics Syst Pharmacol 6:29-39, 2017
31. Feng Y, Wang X, Suryawanshi S, et al: Linking tumor growth dynamics to survival in ipilimumab-treated patients with advanced melanoma using mixture tumor growth dynamic modeling. CPT Pharmacometrics Syst Pharmacol 8:825-834, 2019
32. Benzekry S: Artificial intelligence and mechanistic modeling for clinical decision making in oncology. Clin Pharmacol Ther 108:471-486, 2020
33. Bruno R, Bottino D, de Alwis DP, et al: Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models. Clin Cancer Res 26:1787-1795, 2020
34. Bender BC, Schindler E, Friberg LE: Population pharmacokinetic–pharmacodynamic modelling in oncology: A tool for predicting clinical response. Br J Clin Pharmacol 79:56-71, 2015
35. Fu J, Fang MJ, Dong D, et al: Heterogeneity of metastatic gastrointestinal stromal tumor on texture analysis: DWI texture as potential biomarker of overall survival. Eur J Radiol 125:108825, 2020
36. Bak SH, Park H, Sohn I, et al: Prognostic impact of longitudinal monitoring of radiomic features in patients with advanced non-small cell lung cancer. Sci Rep 9:8730, 2019
37. Wang ZI, Mao LI, Zhou Zg, et al: Pilot study of CT-based radiomics model for early evaluation of response to immunotherapy in patients with metastatic melanoma. Front Oncol 10:1524, 2020