Chapter 1

Data on patterns of lymphatic progression

One critical aspect of our effort to model and predict the lymphatic tumor progression is the data we use to train the model. As previously explained, our model essentially consumes tables with rows of patients and columns involvement by lymph node level (LNL). Data in this relatively simple format has been extracted in the past to create studies like [2] or [6]. However, the authors then used the data to compute statistics of it – e.g. the prevalence of involvement – but stopped short of publishing that data in its raw format. From these statistics it is – with one exception [5] – usually not possible to reconstruct the correlations between the involvement of LNLs.

With almost no usable data, of course, our methodology for modelling lymphatic progression cannot be tested or applied. So, we decided to start at the University Hospital Zurich (USZ) to extract all patterns of lymphatic progression in patients with newly diagnosed oropharyngeal squamous cell carcinoma (OPSCC) between 2013 and 2019. We then not only used that data for inference on it, but also published it freely, hoping that other researchers might find it useful and that it may even motivate them to share their data in a similar fashion in the future.

In the following sections, I will include large parts of the publication [4], in which we detailed the extraction of the dataset, its characteristics and how we made it available. It is important to note that the first authorship is shared in this publication: Jean-Marc Hoffmann, a radiation oncologist at the USZ, extracted most of the data from digital patient and imaging records. Bertrand Pouymayou, a medical physicist and postdoctoral researcher at the USZ built up a complex template for easier extraction and storage of the patient information. He also created the initial interface for viewing the data. My contribution to this work was the processing of the data, creating figures and tables for the publication, host the cohort in the form of a comma separated values (CSV) table in online repositories and, lastly, develop and deploy an online interface akin to what Bertrand Pouymayou had implemented earlier.

1.1 Detailed patient-individual reporting of lymph node involvement in oropharyngeal squamous cell carcinoma with an online interface

1.1.1 Abstract

Purpose/Objective

Whereas the prevalence of LNL involvement in head and neck squamous cell carcinoma (HNSCC) has been reported, the details of lymphatic progression patterns are insufficiently quantified. In this study, we investigate how the risk of metastases in each LNL depends on the involvement of upstream LNLs, T-category, Human Papillomavirus (HPV) status and other risk factors.

Results

We retrospectively analyzed patients with newly diagnosed OPSCC treated at a single institution, resulting in a dataset of 287 patients. For all patients, involvement of LNLs I-VII was recorded individually based on available diagnostic modalities (positron emission tomography (PET), magnetic resonance imaging (MRI), computed tomography (CT), fine needle aspiration (FNA)) together with clinicopathological factors. To analyze the dataset, a web-based graphical user interface (GUI) was developed, which allows querying the number of patients with a certain combination of co-involved LNLs and tumor characteristics.

Results

The full dataset and GUI is part of the publication. Selected findings are: Ipsilateral level IV was involved in 27% of patients with level II and III involvement, but only in 2% of patients with level II but not III involvement. Prevalence of involvement of ipsilateral levels II, III, IV, V was 79%, 34%, 7%, 3% for early T-category patients (T1/T2) and 85%, 50%, 17%, 9% for late T-category (T3/T4), quantifying increasing involvement with T-category. Contralateral levels II, III, IV were involved in 41%, 19%, 4% and 12%, 3%, 2% for tumors with and without midline extension, respectively. T-stage dependence of LNL involvement was more pronounced in HPV negative than positive tumors, but overall involvement was similar. Ipsilateral level VII was involved in 14% and 6% of patients with primary tumors in the tonsil and the base of tongue, respectively.

Conclusions

Detailed quantification of LNL involvement in HNSCC depending on involvement of upstream LNLs and clinicopathological factors may allow for further personalization of elective clinical target volume (CTV-N) definition in the future.

1.2 Material & methods

1.2.1 Data curation

Test

We included patients diagnosed with OPSCC (primary diagnosis) between 2013 and 2019 and treated at the department of radiation oncology and/or head and neck surgery of the USZ. Patients with prior radiotherapy or surgery to the neck were excluded, resulting in a dataset of 287 patients. Specific subsites of oropharyngeal cancer included the base of tongue, the tonsils as well as the oropharyngeal side of the vallecula and the posterior or lateral wall of the oropharynx. Patient information consisted of the date of birth, gender, the date of the 1st histological confirmation of the tumor, the performed treatment (surgery with neck dissection prior to RT/RCHT vs. surgery only vs. definitive radio(chemo)therapy), risk factors such as nicotine abuse and HPV-status (p16 pos/neg), the TNM-classification (UICC 7th edition until 2017, 8th edition since 2017), the position of the primary tumor (left/right neck) as well as positive vs. negative mid-sagittal plane extension. Further details are described in the accompanying data-in-brief article [3].

The analysis of the lymphatic spread included levels Ia, Ib, IIa, IIb, III, IV, V, VII and was performed separately for the diagnostic imaging modalities available for a patient (FDG PET-CT, FDG PET-MRI, MRI, CT) as well as FNA and radiotherapy planning CT if available. This was performed by 2 experienced radiation oncologists by reviewing radiology and pathology reports together with the diagnostic images. Criteria for considering a lymph node as malignant followed the description in Biau et al [1] and are described in detail in the data-in-brief article [3].

1.2.2 Data base