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. Clearly, no model can ever be useful without sufficient and detailed data.

1.1 Material methods

1.1.1 Data curation

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 University Hospital Zurich (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 [**ludwig˙dataset˙2021**].

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 [**biau˙selection˙2019**] and are described in detail in the data-in-brief article [**ludwig˙dataset˙2021**].

1.1.2 Data base