0.1 Material & methods

Data curation

We included patients diagnosed with oropharyngeal squamous cell carcinoma (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 Human Papillomavirus (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 positron emission tomography (PET)-computed tomography (CT), FDG PET-magnetic resonance imaging (MRI), MRI, CT) as well as fine needle aspiration (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].

Data base

The full dataset is available as a comma separated values (CSV)-file via the data-in-brief article linked to this publication [3] and on GitHub at `https://github.com/rmnldwg/lydata` in a folder named `2021-usz-oropharynx`.

In addition, the dataset has been archived and given a persistent identifier: `https://doi.org/10.5281/zenodo.6024778`.

Graphical user interface

We developed an online graphical user interface (GUI) based on the Django framework [2] and provide it to explore the dataset. It allows the user to conveniently determine the number of patients that show a particular combination of co-involved LNLs and tumor characteristics. The GUI is available at `https://2021-oropharynx.lyprox.org`; its source code under MIT license is available on GitHub at `https://github.com/rmnldwg/lyprox`. Documentation is provided within the GUI; a video demonstrating the use of the GUI is available online.