

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 [4] or [19]. 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 [18] – 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 (OP-SCC) 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 [14], 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.

1.1.2 Introduction

HNSCC spread through the lymphatic system of the neck and form metastases in regional lymph nodes. Therefore, the target volume in radiotherapy of HNSCC

patients includes, in addition to the primary tumor, parts of the lymph drainage volume [2], [10]. The nodal gross tumor volume nodal gross tumor volume (GTV-N) contains detectable macroscopic lymph node metastases, while the elective clinical target volume CTV-N contains parts of lymph drainage volume that is at risk of harboring microscopic tumor, i.e. occult metastases that are not yet visible with current imaging techniques.

GTV-N definition is primarily performed through imaging techniques (PET-CT/MRI, MRI or CT) as well as FNA. Imaging criteria for lymph node metastases include size, round rather than oval shape, central necrosis, and FDG uptake as summarized by Biau et al [2]. Goel et al. gives an overview over clinical practice in PET/CT for the management of HNSCC [6]. However, all imaging techniques have finite sensitivity and specificity [16], [12], [17], i.e. they fail to detect small metastases or may incorrectly identify suspicious lymph nodes as tumor.

For standardized reporting of the location of lymph node metastases as well as delineation of the CTV-N, the lymph drainage system of the neck is divided into anatomically defined regions called LNL [8]. CTV-N definition amounts to the decision which LNLs to include into the elective CTV-N and is based on international consensus guidelines. Such guidelines were first published by Grégoire et al in 2000 and have been updated in 2006, 2014 and 2019 [2], [8], [7], [9]. Current recommendations for the selection of lymph node levels in OPSCC can be found in Table 2 of the guidelines published in 2019 by Biau et al. [2]. Current guidelines are primarily based on the prevalence of LNL involvement for a given primary tumor location, i.e. the percentage of patients diagnosed with metastases in each level. It is recommended that the elective CTV-N includes all LNLs that are involved in 10–15% of patients or more. Patients are primarily stratified by primary tumor location. For example, tumors of the soft palate, the posterior pharyngeal wall and the base of tongue show lymph node metastases on both sides via crossing lymph vessels. For this reason, even for lateralized tumors of these localizations, bilateral neck treatment is recommended. However, the lymphatic drainage of the tonsil is mainly unilateral, therefore an ipsilateral irradiation is recommended for lateralized low T-category (T1/T2) tumors (at least up to lymph node stage N2a). Volume-reduced elective nodal irradiation has been or is being investigated in several trials [3], [15].

While the general patterns of lymph drainage in the neck is understood and prevalence of LNL involvement has been reported in the literature [7], [4], [11], [1], the details of progression patterns in OPSCC are poorly quantified. How much does the risk of level IV involvement increase depending on whether levels II and III harbors macroscopic metastases? How much does the risk of involvement increase for late versus early T-category? Are progression patterns different for HPV positive versus HPV negative tumors? Answering these questions quantitatively may allow for further personalizing CTV-N definition based on an individual patient's clinical presentation at the time of diagnosis.

The basis for better quantification of LNL involvement are detailed datasets of HNSCC patients for whom involvement is reported per individual LNL together with tumor and patient characteristics. For example, to answer the question of how much the risk in level IV increases depending on the involvement of upstream levels II and III, it is insufficient to only report the prevalence of LNL involvement in levels II, III, and IV. Instead, the observed frequency of certain involvement

combinations must be known, e.g. how often levels II, III and IV are involved simultaneously, versus how often only the levels II and III are involved without level IV. The contributions of this work are:

- We provide a dataset of lymphatic progression patterns in 287 OPSCC patients treated at our institution in whom involvement of LNLs together with tumor characteristics are reported on a patient-individual basis.
- To visualize and explore the complex dataset, a graphical user interface is provided that allows the user to query the number of patients who were diagnosed with a specific combination of simultaneously involved LNLs and tumor characteristics.

We hope that this work provides the basis for collecting large multicenter datasets of lymphatic progression patterns, which can then inform future guidelines on further personalized CTV-N definition.

1.1.3 Material & methods

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 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 [13].

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 [2] and are described in detail in the data-in-brief article [13].

1.1.4 Data base

The full dataset is available as a CSV-file via the data-in-brief article linked to this publication [13] and on GitHub at <https://github.com/rmnlldwg/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>.

1.1.5 Graphical user interface

We developed an online GUI based on the Django framework [5] 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 in the supplementary materials.