IntechOpen *Journals*

Digital Medicine and Healthcare Technology

RESEARCH PAPER

OpenAi's ChatGPT-4, BARD and YOU.com (AI) and the Cancer Patient, for Now, Caveat Emptor, but Stay Tuned

Glenn Tisman[1,*] and Raju Seetharam[2]

1 Independent Medical Oncology Investigator, Florida, USA; Alumnus, Doctor of Medicine, Buffalo School of Medicine, New York, USA
2 Independent Software Scientist, Florida, USA; Alumnus, Masters in Computer Science, Rochester Institute of Technology, USA
*Corresponding author. E-mail: glennmd@gmail.com

Abstract

ChatGPT-4, BARD, and YOU.com are AI large language models (LLM) developed by OpenAI based on the GPT-3-4 architecture and Google. They were trained using unsupervised learning, which allows them to learn from vast amounts of text data without requiring explicit human labels. ChatGPT-4 was exposed to training information up to September 2021. By presenting prompts (queries) to ChatGPT-4, BARD, and YOU.com, including a typical case presentation (vignette) of a new patient with squamous cell tonsillar cancer, we uncovered several specific issues that raise concerns for the current application of this early phase of advanced LLM AI technology for clinical medicine. By prompting and comparing responses of three different LLMs (ChatGPT-4, BARD, and YOU.com) to identical prompts, we reveal several flaws in each AI that, if taken as factual, would affect clinical therapeutic suggestions and possible survival. The presented clinical vignette of a patient with newly diagnosed tonsillar cancer is presented to three LLMs readily available for free trial allowing comparison of results. We observed frequent changing responses to unchanging prompts over just hours and days within the same and between LLMs, critical errors of guideline-recommended drug therapy, and noted that several AI-supplied references presented by the AIs are bogus AI-generated references whose DOI and or PMID identifiers were either nonexistent or led to completely irrelevant manuscripts on other subjects.

Keywords: AI, ChatGPT-4, BARD, YOU.com, OpenAI, neural, tonsillar, cancer

Citation
Glenn Tisman and Raju Seetharam (2023), OpenAi's ChatGPT-4, BARD and YOU.com (AI) and the Cancer Patient, for Now, Caveat Emptor, but Stay Tuned. *Digital Medicine and Healthcare Technology* 2(1), 1–27.

DOI
https://doi.org/10.5772/dmht.19

Published
25 September 2023

1. Introduction

There have been many so-called AI winters in the past several years. These are periods when the hype of artificial intelligence (AI) (or augmented intelligence to some) has fizzled in disappointing failure. Presumably, we are now in a newer and

deeper learning period for AI. Newscasters have put the world on fire with excitement and explosive, if not irrational, exuberance for OpenAi's ChatGPT [1–3]. OpenAI claims human-level performance in academic and professional settings, which includes passing the bar exam [4] with scores in the top 10% of test takers. GPT-4 is touted to have improved performance in understanding the nuances of language, including its content, tone, and meaning. We, therefore, resolved to evaluate the accuracy of medical information gleaned from not only ChatGPT-4 but also two of its competitors, BARD and YOU.com. We understand from many sources generalizations stating that ChatGPT-4, though universally available for a free trial, should not be used for clinical medical purposes at this time. However, AI is currently being used to help interpret mammograms [5], chest CTs [6], chest X-rays [7], pathology slides [8] and the like, so we investigated using ChatGPT-4, BARD, and YOU.com AIs for clinical decision-making for an accurate clinical vignette of a patient with a new diagnosis of squamous cell cancer of the tonsil. Instead of generalizations, we attempted to uncover real-life and detailed weaknesses of ChatGPT-4 and other AIs and included the available Large Language Models, AIs, BARD, and YOU.com for comparison. We anticipated that though most current AIs decline to offer patients specific medical advice, they do suggest therapeutic solutions to users identified as physicians. This manuscript will take a deep dive into a new case of tonsillar cancer predicated on the necessity of knowing the patient's clinical tumor staging results. For those wanting to familiarize themselves with general therapeutic principles of head and neck cancers, including squamous cell cancer of the tonsil, we provide an excellent general review of the subject [9].

1.1. Note related to HPV-positive tonsillar cancer

The oral mucosa is exposed to the infectious human papillomavirus (HPV) generally from vaginal-oral sex (22.4%), anal-oral sex (4.4%), excessive alcohol (3.5%), and unknown causes (70%). Viral persistence can lead to a progression to invasive squamous cell cancer of the oropharynx. Persistent HPV infections may progress to invasive cancer within ten years. It is estimated that 60% of all oropharyngeal cancers are caused by HPV infections. The most common high-risk HPV types are 16 and 18, but also include the less prevalent 31, 33, 45, 52, and 58 types However, most infections are naturally cleared in 12 to 24 months. The incidence of HPV-associated oropharyngeal squamous cell carcinoma (OPSCC) is quickly increasing and rarely discussed among the general population. In the United States, the incidence of HPV-positive OPSCC was 4.62 per 100,000 persons, and it now represents the sixth most common incidence of nonskin cancer among white male patients younger than 65 years. HPV-associated tonsillar tumors carry a much better prognosis than tumors negative for the virus. In the United States, about 80% of oropharyngeal cancer cases in men and about 65% of cases in women are attributable to smoking and tobacco use.

2. *Materials and methods: using ChatGPT-4, BARD, and YOU.com*

To corroborate inaccurate results from the three AI sources, general Google searches and/or procurement of published manuscripts were used for verification purposes.

ChatGPT-4 was accessed by simply entering the URL into the preferred browser or selecting the link: https://openai.com/product/gpt-4 and following the instructions. The user submits text prompts or queries into the ChatGPT-4 dialogue box, and it returns textual answers to prompts.

Bard is Google's conversational AI https://bard.google.com Google's experimental, conversational AI chat service. It is meant to function similarly to ChatGPT, with the most significant difference being that BARD will pull its information from the web. Med PALM2 is a large language model (LLM) developed by Google. It is a medical version of PALM2, which is a general-purpose LLM. Med PALM2 is currently being trained on a massive dataset of medical text and code, and it can generate text, translate languages, write different kinds of creative content, and answer medical questions in a relatively safe and intellectually informative way. It is currently available only to a select group of users. It may represent the future of medical specialty AI.

You.com is a search engine that provides personalized search results based on users' interests and preferences. It uses advanced artificial intelligence and machine learning algorithms to provide a more human-like and intuitive search experience. It is accessible at YOU.com, and we have found that it has performed very well in responding to medical prompts.

2.1. ChatGPT-4 explains how it works

Before starting, we asked ChatGPT-4 through the following prompt (MY PROMPT 1) to explain itself.

> ***MY PROMPT 1*** (*the first query* submitted to ChatGPT-4 follows):
>
> Where can I find a simple summary explanation of how ChatGPT-4 works?
>
> **ChatGPT-4** returned the following textual response to MY PROMPT 1:
>
> *A simple summary of how ChatGPT-4 works can be found below:*
>
> *ChatGPT-4 is an AI language model developed by OpenAI based on the GPT-4 architecture. It is trained using unsupervised learning, which allows it to learn from vast amounts of text data without requiring explicit human labels.*

The model works in two main steps: pre-training and fine-tuning. In the pre-training phase, ChatGPT-4 learns to predict the next word in a sentence from a large dataset containing parts of the internet. This helps it understand grammar, facts about the world, and some reasoning abilities. However, it might also learn some biases present in the training data.

In the fine-tuning phase, the model is refined on a narrower dataset with specific tasks and goals, often generated with the help of human reviewers. This phase helps ChatGPT-4 understand the context and respond more accurately to user inputs.

ChatGPT-4 uses a deep neural network called a Transformer, which has a large number of layers and attention mechanisms. This enables it to process and generate human-like text effectively. When given a text prompt or a question, ChatGPT-4 generates a response by predicting the most likely sequence of words based on the knowledge it has acquired during training.

Recalling that ChatGPT-4 might generate plausible-sounding but incorrect or nonsensical answers (RS interjects termed neural net hallucinations by some software experts) (10), as its primary goal is to predict text and not to reason deeply. ***It is essential to verify the information provided by the model and treat it as a starting point rather than a definitive answer****. We have also noted that, at times, its responses may simulate what psychiatrists term confabulation or repeating nonsense without malcontent.*

END

We then tested ChatGPT-4, BARD, and YOU.com with submitted prompts (some identical) to the different AIs and evaluated the accuracy and possible consequences of inaccurate and inappropriate responses. These prompts had been pretested with several similar textual versions of prompts that appeared to get a comprehensive (but rarely accurate) response from all of the AIs.

3. Results

The following is a representative sampling of the variation of prompt responses concentrating on ChatGPT-4.

3.1. Evaluating a sample of AI prompt responses seeking correct tonsillar clinical staging for our patient presented in the clinical vignette

MY PROMPT 2-A submitted to ChatGPT-4

Could you stage the following tonsillar tumor? A 64-year-old Caucasian gentleman had noted a fullness in the left throat and consulted an ENT

physician. CT of the neck with contrast revealed a slight interval increase in the size of the enhancing lesion involving the left palatine tonsil from the prior CT, consistent with the patient's known squamous cell carcinoma of the left tonsil. There was a slight interval increase in the size of the enlarged left metastatic level IIA cervical lymph node containing solid and cystic components. This node abuts the posterior margin of the left submandibular gland. CT with the contrast of the chest, abdomen, and pelvis revealed the absence of metastatic disease. The patient underwent a transoral robotic resection (TORS) of the oropharyngeal tonsillar lesion and a left neck dissection. Pathology revealed a 2.7 cm left tonsillar squamous cell carcinoma absent special features. Notably, the tumor was HPV-associated, and specimen margins were negative for tumor; however, there were close margins: 1–2 mm (deep lateral margin at the inferior aspect of tumor at 2 mm (inferior mucosa), 3 mm (deep inferomedial) and there was an additional 3–4 mm close margin. There was an absence of lymphovascular or perineural invasion. A level II–IV left cervical lymph node dissection revealed 1 of 32 removed lymph nodes involved with the tumor. The abnormal node was at level II, 4 cm in size, and contained SCC without extranodal extension.

END

ChatGPT-4 returned the following textual response to MY PROMPT 2-A:

Based on the provided information, we can stage the tonsillar tumor using the TNM (Tumor, Node, Metastasis) classification system by the American Joint Committee on Cancer (AJCC).

Tumor (T): The primary tumor is 2.7 cm in size and was in the left palatine tonsil. According to the AJCC 8th Edition, this would be classified as T2 (T2: tumor >2 cm but ≤4 cm in greatest dimension).

Node (N): There is one involved lymph node at level II, which is 4 cm in size and contains squamous cell carcinoma (SCC) without extranodal extension (ENE). This would be classified as N1 (N1: Metastasis in a single ipsilateral lymph node, ≤6 cm in greatest dimension, ENE (–)). (Here, ChatGPT-4 uses the AJCC 8th Ed system, that is HPV-positive to define N1 GT).

Metastasis (M): CT scans with contrast of the chest, abdomen, and pelvis revealed an absence of metastatic disease. This would be classified as M0 (M0: No distant metastasis).

So, the overall stage of the tonsillar tumor based on the TNM classification system would be ***Stage IVA (T2 N1 M0)***.

END

WRONG

The correct answer is Clinical Stage I (T2N1M0)

MY PROMPT 2-B

Are you sure this is an AJCC 8th ED TNM Stage IVA tonsillar tumor?

ChatGPT-4 Response to 2-B

I apologize for the confusion in my previous response. Upon closer evaluation, the correct staging of the tonsillar tumor should be:

Tumor (T): T2 (T2: tumor >2 cm but ≤4 cm in greatest dimension)

Node (N): N1 (N1: Metastasis in a single ipsilateral lymph node, ≤6 cm in greatest dimension, ENE (–)) Metastasis (M): M0 (M0: No distant metastasis)

According to the AJCC 8th Edition for Oropharyngeal Cancer (p16-positive), the correct overall stage of this tonsillar tumor should be ***Stage II (T2 N1 M0) and not Stage IVA*** *as previously mentioned. I apologize for the confusion, and I hope this clears up any misunderstandings.*

END

WRONG

Correct answer is Clinical Stage I (T2N1M0)

MY PROMPT 3-A This prompt is identical to the prior MY PROMPT 2-A (see above)

ChatGPT-4 RESPONSE to PROMPT 3-B (This is the correct response for staging)

Based on the information you provided and using the American Joint Committee on Cancer (AJCC) 8th Edition TNM staging system for oropharyngeal squamous cell carcinoma (as of September 2021, which is when my knowledge was last updated), we can break down the tumor staging as follows: T (Tumor): The tumor size is 2.7 cm and is in the left palatine tonsil. It is mentioned that the specimen margins were negative for the tumor. There's no mention of the tumor invading nearby structures beyond the tonsil. Given this information, it could be classified as T2 (a tumor more than 2 cm but not more than 4 cm in greatest dimension).

N (Nodes): One of 32 lymph nodes is involved with the tumor, and this lymph node is located at level II and measures 4 cm in size. This might be classified as N2b (a single ipsilateral node larger than 3 cm but not larger than 6 cm).

M (Metastasis): The information states that CT scans of the chest, abdomen, and pelvis revealed the absence of metastatic disease, which could classify as M0 (no distant metastasis). ***Since the tumor is HPV-associated****, this can also impact staging. The AJCC 8th edition has separate staging criteria for HPV-positive oropharyngeal tumors compared to HPV-negative tumors. Putting this together, the tumor might be classified as T2 N2b M0 according to the AJCC 8th edition. This could correspond to:* ***Stage I for HPV-positive oropharyngeal cancer or Stage III/IVA for HPV-negative oropharyngeal cancer****, depending on additional factors.*

However, as mentioned before, it is important to recognize that I am an AI and not capable of making clinical decisions or diagnoses. Staging of cancer should be conducted by a qualified healthcare professional who can evaluate all of the clinical data. Also, the TNM staging criteria may undergo updates, so it's important to consult the latest guidelines.

END

Correct for clinical stage I; however, N stage is wrong

The given N description is incorrect. It should be N1, not N2b

Interestingly, the Correct answer is Clinical Stage I for HPV+; however correct N designation is (T2 N1 M0), not N2b.

ChatGPT-4 MYPROMPT 2-C, which is identical to MYPROMPT 2-A and MYPROMPT 2-B

MY PROMPT RESPONSE 2-C

Based on this information, the T category could likely be T1 (tumor ≤4 cm in most significant dimension).

N (Nodes): One of 32 lymph nodes is involved with the tumor, and this lymph node is located at level II and is 4 cm in size. This might be classified as N2b (single ipsilateral node larger than 3 cm but not larger than 6 cm).
M (Metastasis): The information states that there is no distant metastasis, which would classify as M0. Using the TNM categories and considering the tumor is HPV-positive, this tumor could be classified as T1N2bM0.
According to the AJCC 8th Edition for HPV-positive oropharyngeal cancers, this ***corresponds to Stage II****.*

Please note that the clinical staging of a tumor should be performed by a qualified healthcare professional who has access to all the relevant data and can interpret it in the context of the most up-to-date guidelines and the patient's survival clinical picture. Always consult a medical professional for definitive cancer staging and treatment recommendations.

END

WRONG

Correct answer is Clinical Stage I (T2N1M0)

MY PROMPT 3

Are you absolutely sure this is an AJCC 8th ED TNM Stage IVA tonsillar tumor?

ChatGPT-4 Response 3

I apologize for the confusion in my previous response. Upon closer evaluation, the correct staging of the tonsillar tumor should be:

Tumor (T): T2 (T2: tumor >2 cm but ≤4 cm in greatest dimension)

Node (N): N1 (N1: Metastasis in a single ipsilateral lymph node, ≤6 cm in greatest dimension, ENE (–)) Metastasis (M): M0 (M0: No distant metastasis)

According to the AJCC 8th Edition for Oropharyngeal Cancer (p16-positive), the correct overall stage of this tonsillar tumor should be ***Stage II (T2 N1 M0)*** *and not Stage IVA as previously mentioned. I apologize for the confusion, and I hope this clears up any misunderstandings.*

END

WRONG

Correct answer is Clinical Stage I (T2N1M0)

MY PROMPT 4

Let's try again; I wonder whether this tumor is Stage I or Stage II.

ChatGPT-4 Response

I apologize for the confusion in my previous responses. Based on the provided information and considering the tumor is HPV-associated (p16-positive), the correct staging according to the AJCC 8th Edition for Oropharyngeal Cancer should be:

Tumor (T): T2 (T2: tumor >2 cm but ≤4 cm in greatest dimension)

Node (N): N1 (N1: Metastasis in a single ipsilateral lymph node, ≤6 cm in greatest dimension, ENE (–))

Metastasis (M): M0 (M0: No distant metastasis)

The correct overall stage of the tonsillar tumor for HPV-associated oropharyngeal cancer should be ***Stage I (T2 N1 M0).*** *The staging system for HPV-associated oropharyngeal cancer differs from the staging for*

non-HPV-associated cancers, which is why my previous responses were incorrect. I apologize for the confusion and any inconvenience this may have caused.

END

Correct answer is Clinical Stage I (T2N1M0)

Considering BARD PROMPTS and its responses

BARD PROMPT

Could you stage the following tonsillar tumor? A 64-year-old Caucasian gentleman had noted a fullness in the left throat and consulted an ENT physician. CT of the neck with contrast revealed a slight interval increase in the size of the enhancing lesion involving the left palatine tonsil from the prior CT, consistent with the patient's known squamous cell carcinoma of the left tonsil. There was a slight interval increase in the size of the enlarged left metastatic level IIA cervical lymph node containing solid and cystic components. This node abuts the posterior margin of the left submandibular gland. CT with contrast of the chest, abdomen, and pelvis revealed the absence of metastatic disease. The patient underwent a transoral robotic resection (TORS) of the oropharyngeal tonsillar lesion and a left neck dissection. Pathology revealed a 2.7 cm left tonsillar squamous cell carcinoma absent special features. Notably, the tumor was HPV-associated, and specimen margins were negative for tumor; however, there were close margins: 1–2 mm (deep lateral margin at the inferior aspect of tumor at 2 mm (inferior mucosa), 3 mm (deep inferomedial) and there was an additional 3–4 mm close margin. There was an absence of lymphovascular or perineural invasion. A level II–IV left cervical lymph node dissection revealed 1 of 32 removed lymph nodes involved with the tumor. The abnormal node was at level II, 4 cm in size, and contained SCC without extranodal extension.

BARD RESPONSE

Sure, I can help you stage the tonsillar tumor.

The patient's tumor is 2.7 cm in size, and there is one positive lymph node that is 4 cm in size. The lymph node does not have any extranodal extension. The patient does not have any distant metastasis.

The TNM staging system is used to stage cancer. The TNM system classifies cancer based on the size of the tumor (T), the extent of lymph node involvement (N), and the presence of distant metastasis (M).

In this case, the patient's tumor is T2 (2.7 cm). There is one positive lymph node, which is classified as N1a. There is no distant metastasis, so the patient's stage is M0.

The palatine's stage is, therefore, T2N1aM0, which is a stage IIA cancer.

The patient's tumor is HPV-associated, which means that it is caused by the human papillomavirus. HPV-associated tumors are typically less aggressive than tumors that are not HPV-associated.

The patient's margins were close, which means that there was a small amount of tumor that was not removed with the surgery. This can increase the risk of recurrence.

The patient's lymph node dissection was negative for other involved lymph nodes, which is good news. This means that the cancer has not spread to other lymph nodes in the neck.

The patient's survival prognosis is good. Stage IIA cancers are typically very treatable, and the patient's tumor is HPV-associated, which also makes it less aggressive.

However, the patient's close margins increase the risk of recurrence, so the patient will need to be monitored closely for any signs of cancer returning.

END

WRONG

Correct answer is Clinical Stage I (T2N1M0)

YOU.com

YOU.com PROMPT Same prompt as given to BARD above.

YOU.com RESPONSE

Based on the information provided, the tonsillar tumor would be staged as follows:

T: T2 (size of 2.7 cm) N: N1 (involvement of one lymph node at level II) M: M0 (absence of metastatic disease)

Overall stage: Stage II (based on the TNM staging system)

END

WRONG

Correct answer is Clinical Stage I (T2N1M0)

The same MYPROMPT 2A was fed to BARD and YOU.com, and they both answered stage II. It is recalled that the correct answer is stage I.

ChatGPT-4 initially responded with three incorrect, followed by two correct staging evaluations (Stage IVA/III: IVA: II: I: I). It appears that responses to the same prompt varied even though the AJCC 8th Edition has been around since 2017–2018, and that edition defined the newest staging criteria for tonsillar cancer based on tumor HPV+ or HPV- status.

ChatGPT-4*	**Response**	**Status** *(see Appendix for correct staging criteria)*
First, try	Stage IVA	Wrong
Second try	Stage III/IVA	Wrong
Third try	Stage II	Wrong
Fourth try	Stage I	Correct
Fifth try	Stage I	Correct
BARD try	Stage II	Wrong
YOU.com try	Stage II	Wrong

*NOTE: the above variation of responses depended on the day and even the hour the same prompt was fed to ChatGPT-4 and, in some instances, the follow-up questioning/prompt of ChatGPT-4 after reading an incorrect response. Some have claimed that LLMs can hallucinate [10]. Both BARD and YOU.com were correct for the nodal stage of N1 but incorrect for the final clinical stage.

When we rely on today's LLM AI in medicine, in some respect, we are dealing with a multitude of unknowns within a black box, and it is close to impossible for users to apply logical reason(s) for the discrepancies and inaccuracies delivered by LLM AIs. After all, AJCC's 7th and 8th editions were fully available for AI access during their essential training period years before our prompts were submitted. The staging errors are unacceptable because all therapeutic recommendations are stage related. Were such gross errors a manifestation of what some AI scientists call a neural net hallucination [10], a sophisticated apology for a wrong answer?

The true AJCC 8th ED [11] stage for this patient's tumor is **Stage I**, not **Stage IVA/III**, not **Stage II**, and it took multiple wrong guesses by ChatGPT-4 to arrive at the correct clinical tumor stage. BARD and YOU.com were tested less but consistently came up with the wrong answer. In one illustrated prompt, the following prompt appeared to "coax or cast doubt" to ChatGPT-4 (see MYPROMPT 4), thus seemingly encouraging it "to produce other answers", and it did such. ChatGPT-4 remembered the history of its previous but incorrect responses (Stage IVA and Stage II)—the physician's additional probing prompts. The follow-up prompt was submitted because the physician was aware of the correct answer.

We conjecture that ChatGPT-4 may have partially ignored or misinterpreted our initial vignette, MY PROMPT 2-A, revealing that the patient's tumor was HPV-associated, but we are unsure. The mechanistic complexity of multilayered neural network "thinking" must be left to the experts [12].

4. Evaluating the potentially dangerous consequences of AI initial incorrect clinical staging responses

Tonsillar squamous cell cancer is a successfully treatable disease when diagnosed early and associated with the human papillomavirus (HPV positivity). However, an advanced-stage illness, especially in the presence of HPV-negativity, continues to have a poor prognosis. AJCC 8th Ed Stage IV disease that is HPV-negative carries a 5-year survival of only 50% or less, while 5-year survival of HPV-positive tonsillar cancer reaches 65% [13]. In addition to the poor prognosis of advanced-Stage IV diseases, treatment of such high stages is associated with increased patient morbidity. Toxic therapeutic effects may include sore, dry mouth, dental issues, difficulty chewing and swallowing, excessive mucous accumulation, voice changes, coughing/choking, etc. [14]. Surgery and chemoradiotherapy alike can lead to significant patient impairment necessitating reliance on tracheostomy and gastrostomy tube placement to prevent aspiration and maintain nutrition. This can have a pronounced negative impact on a patient's quality of life and, for some, may not be reversible [14, 15] Figures 4–8.

Recall that several prompt submissions were required for ChatGPT-4 to present the correct AJCC 8th Ed **Stage I** classification for our patient. Had the treating physicians acted on the wrong staging, that could have resulted in inappropriate therapeutic recommendations, including presurgical chemotherapy [16] and more intensive radiation [17].

Figure 2A and B reveal that the 5-year survival for Stage IV (p16-) tonsillar cancer as staged by the AJCC 7th Ed or 8th Ed is approximately 58%. Sharing such inaccurate staging with a patient could precipitate undue high anxiety.

It is essential to understand that ChatGPT, BARD, and YOU.com are types of large language model (LLM). ChatGPT is based on OpenAI's GPT-3 language model and was trained on a massive amount of data to generate human-like text in response to prompts and only trained to guess words and does not contain any knowledge about medical or any other domains. The LLM is used to power its AI chatbots and to interpret natural language queries, allowing for a more human-like and intuitive search experience. Another AI, i.e., Med-PaLM 2, is a large language model (LLM) developed by Google that is designed to provide high-quality answers to medical questions. It is the second iteration and has been developed with a focus on safety and avoiding biases. Med-PaLM 2 has been trained by Google's health research teams with medical knowledge and provides answers to a variety of complex medical issues. It is an artificial intelligence AI system that is designed to give healthcare professionals access to pretested-accurate medical information. The model's access is limited to a select group of Google Cloud customers.

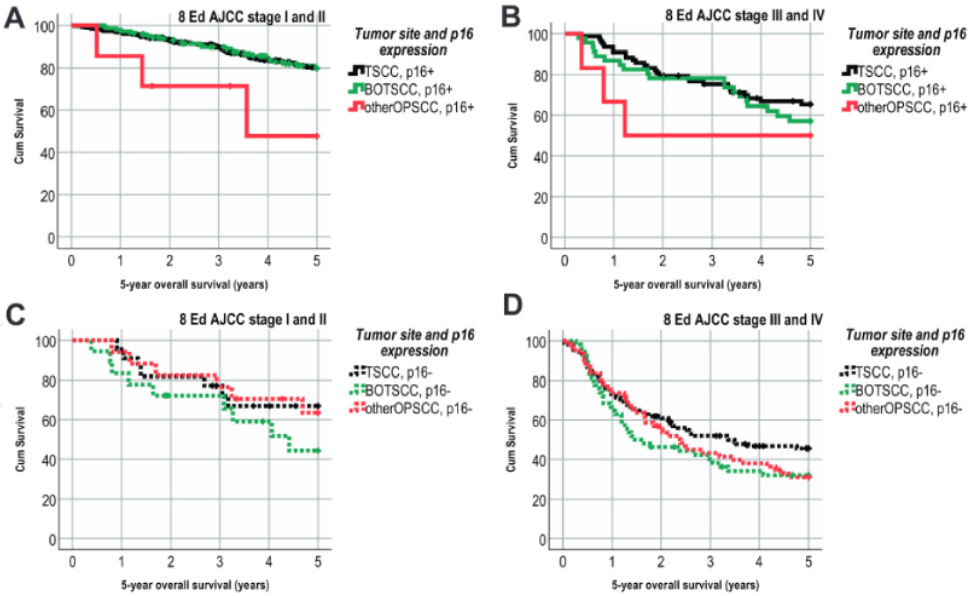

Figure 1. The 5-year cumulative survival for oropharyngeal cancer patients according to the AJCC TNM 8th Ed staging manual. Authors label the worst surviving oropharyngeal tumors as otherOPSCC (otherOPSCC = other than the base of tongue or tonsil as primary) [13]. **See Appendix for AJCC 7th and 8th Ed oropharyngeal staging.**

Had ChatGPT-4 applied the downstaging modulation of HPV-positivity to the incorrect stage IVA, it would have improved the 5-year clinical prognosis to approximately 78% [13] casting more confusion into the process (see Figure 2D). Had we taken ChatGPT-4's original output as factual, our therapeutic assumptions would be erroneous since we would be building on an incorrect **Stage IV** tonsillar cancer base.

Acceptance of other ChatGPT-4 wrong responses, including those made by BARD and YOU.com of **Stage II**, as the appropriate stage for the patient's HPV-positive tumor, would place the patient's 5-year survival at approximately 80%, much better (Figure 1A) but still the wrong stage choice.

Considering ChatGPT-4's correct response, **Stage I**.

In Figure 1A, the 5-year survival of actual Stage I HPV-positive or (p16+) patients would approximate that of Stage II (p16+) patients; however, in the AJCC 8th Ed, there was a statistically significant difference in the 5-year overall survival for stage I and Stage II disease generally varying in different analyses [18, 19]. As an example, van Gysen reported the difference to be between (96.9% for Stage I vs. 77.1% for Stage II respectively; $p < 0.0001$) survival at five years, but not between Stage II

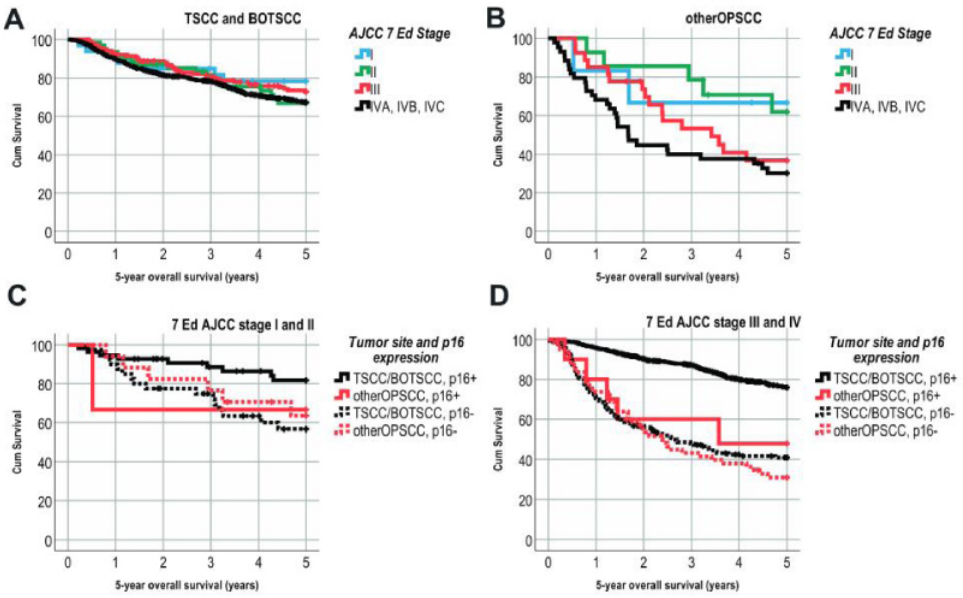

Figure 2. The 5-year cumulative survival for oropharyngeal cancer patients. Note that stage IVA uses the older AJCC 7th Ed for oropharyngeal survival curves [13].

and III disease (p = 0.98) [20]. Classifying the patient's stage correctly as **Stage I** rather than the original **Stage IVA** was a critical ChatGPT-4 error!

Overall survival of patients with oropharyngeal squamous cell carcinoma (OPSCC) per TNM-8 stage categories (I–II and III–IV) according to stratification based on tumor p16 expression status and tumor origin (tonsillar squamous cell carcinoma, TSCC; the base of tongue carcinoma, BOTSCC; non-tonsillar, non-base of tongue carcinoma, otherOPSCC). (1A) Patients with HPV-mediated (p16 overexpressing, p16+) TSCC and BOTSCC TNM-8 stage I–II had significantly better overall survival than patients with HPV-mediated (p16+) otherOPSCC in stage I–II (TSCC/BOTSCC vs. otherOPSCC: log-rank-test: p Z 0.02). (1B) No significant differences between patients with p16+ TSCC/BOTSCC vs. otherOPSCC in stage III–IV were observed (log-rank test: p Z 0.31). (2C-D) No differences in survival between patients with HPV-unrelated (no p16 overexpression, p16-) TSCC/BOTSCC and otherOPSCC, independent of in TNM8 stage I–II (log-rank test: 0.58) or stage III–IV (log-rank test: 0.43). European Journal of Cancer 139 (2020) 192–200. Linda Marklund et al.

Also, see the Appendix for AJCC 7th Ed and 8th Ed oropharyngeal staging.

Overall survival of patients with oropharyngeal squamous cell carcinoma (OPSCC) per subsite (tonsillar squamous cell carcinoma, TSCC; base of tongue carcinoma, BOTSCC; non-tonsillar, non-base of tongue carcinoma, otherOPSCC) according to stratification based on tumor TNM-7 stage categories (I–IV). (2A) No hazard discrimination utilizing the former TNM-7 staging system in patients with

TSCC/BOTSCC (stage I–II vs. III–IV: log-rank test: p Z 0.56). (2B) Significant hazard discrimination utilizing the former TNM-7 staging system in patients with otherOPSCC (stage I–II vs. III–IV: log-rank test: p Z 0.019). (2C) No survival differences between patients with TNM-7 stage I–II and p16+ and p16-otherOPSCC or p16+ otherOPSCC and p16- TSCC/BOTSCC were observed (log-rank test: p Z 0.78 and p Z 1.0 respectively). However, patients with p16+ TSCC/BOTSCC and low TNM-7 stage had significantly better survival than patients with p16- TSCC/BOTSCC and similar stage (p Z 0.010). (2D) No survival differences between patients with TNM-7 stage III–IV and p16+ and p16-otherOPSCC or p16+ otherOPSCC and p16- TSCC/BOTSCC were observed (log-rank test: p Z 0.47 and p Z 0.65 respectively). Patients with p16+ TSCC/BOTSCC and high TNM-7 stage had significantly better survival than patients with p16- TSCC/BOTSCC and similar stage (p < 0.001). European Journal of Cancer 139 (2020) 192e200. Linda Marklund et al.

5. Using the AJCC 8th Ed staging manual, the correct stage for this patient is T2N1M0, or the accurate prognostic clinical stage I (p-16+ or HPV-positive). Now, how should we treat this patient? What therapy is best?

What if we acted upon the initial incorrect Stage IVA suggestion and the accompanying decreased 5-year survival suggested by ChatGPT-4? As revealed in Figure 4, radiation, chemotherapy, and possibly immunotherapy [21, 22] might have been recommended with more significant associated risks of acute and perhaps chronic toxicity for this patient.

Notably, a relatively new study by Yoshida *et al.* [23] entitled Stage I HPV-positive Oropharyngeal Cancer: Should All Patients Receive Similar Treatments?" found that concurrent chemoradiotherapy was associated with improved overall survival only for patients with lymph node-positive, but not lymph node-negative, AJCC 8th Ed **Stage I** HPV-positive oropharyngeal SCC. The authors found no enhanced survival advantage to treating AJCC 8th HPV-positive **Stage I T2N0M0** patients with the addition of cisplatin chemotherapy to radiation therapy (RT). Patients with **N1 Stage I** disease undergoing concomitant chemotherapy were observed to have a 22% decreased relative risk of mortality, Figure 3. Knowing the patient's true stage suggests that Yoshida's work is a reasonable approach for this patient.

> NOTE: As a supplement for those not routinely involved in therapy for oropharyngeal cancer, we include a display of the latest NCCN Guidelines Version 1.2023 for Cancer of the Oropharynx (p16 [HPV]-positive). Figures 4–8 reveal how treatment intensity which is generally related to higher stages of tonsillar cancer, varies with progressive clinical stages [24, 25].

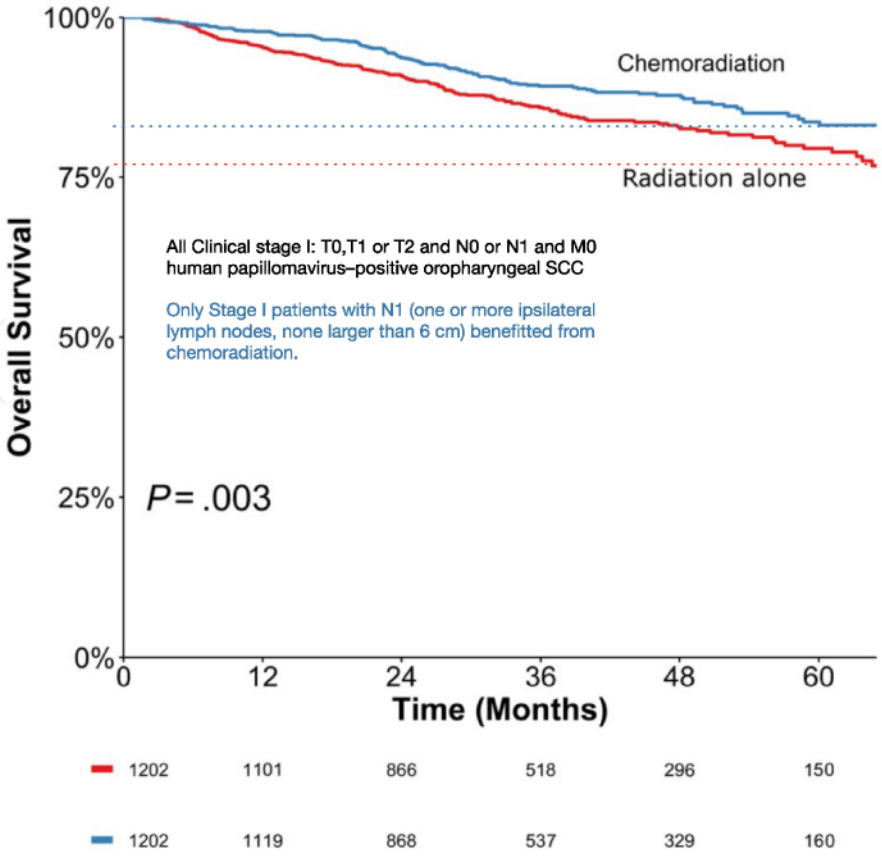

Figure 3. Overall survival curves for chemoradiation vs. radiation alone stage I HPV-positive.

Kaplan-Meier estimates of overall survival among propensity score-matched patients with American Joint Committee on Cancer eighth edition clinical stage I, HPV–positive oropharyngeal squamous cell carcinoma who were treated with concurrent chemoradiation versus radiation alone. Modified from Emi J. Yoshida et al. Stage I HPV-Positive Oropharyngeal Cancer: Should All Patients Receive Similar Treatments? Cancer 2020; 126:58–66.

6. Discussion and conclusions

Accurate staging is an important aspect of cancer care. Careful assessment by imaging, surgical anatomy, pertinent biomarkers (p16±), and the surgical and pathology findings of patients harboring the cancer is crucial to understanding prognosis [18] and making appropriate treatment recommendations.

We have evaluated three AIs' responses to prompts or queries asking for help delineating our patient's tumor stage and observed ChatGPT-4's comments. MY PROMPT 2 presented our patient's clinical vignette and asked ChatGPT-4 to return to the patient's clinical stage. It first responded to Stage IVA, then Stage II, and

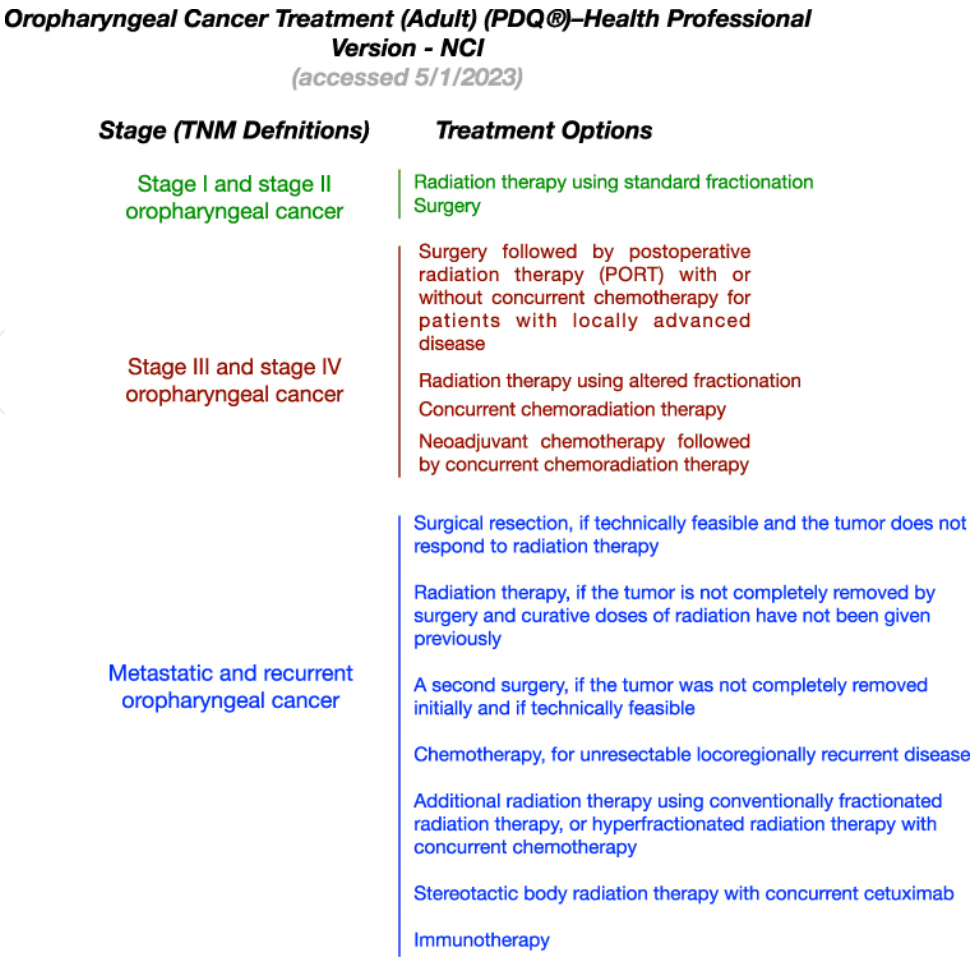

Figure 4. General oropharyngeal treatment recommendations by AJCC 8th Ed TNM staging manual.

finally with the correct answer, Stage I, but only after several tries. The other two AIs agreed on the incorrect stage II answer.

As can be seen in Figures 1–3, each clinical stage is associated with a differing anticipated 5-year survival, and each stage generally implicates different therapeutic modalities and associated toxicities Figures 4–8. In addition, therapeutic decisions are frequently nuanced by intermediate factors, such as histopathological features, including the status of tumor margins and the presence of lymphovascular and perineural invasion [26] and extranodal extension (ENE positive or negative) [27] and social and other coexistent health factors. We had planned to query ChatGPT-4 about those factors; however, we stopped short when we saw it could not accurately evaluate the patient's primary clinical stage with confidence.

At first blush, it appears that the ChatGPT-4, BARD, and YOU.com LLMs missed their mark for this patient and may not have been trained on enough pertinent data.

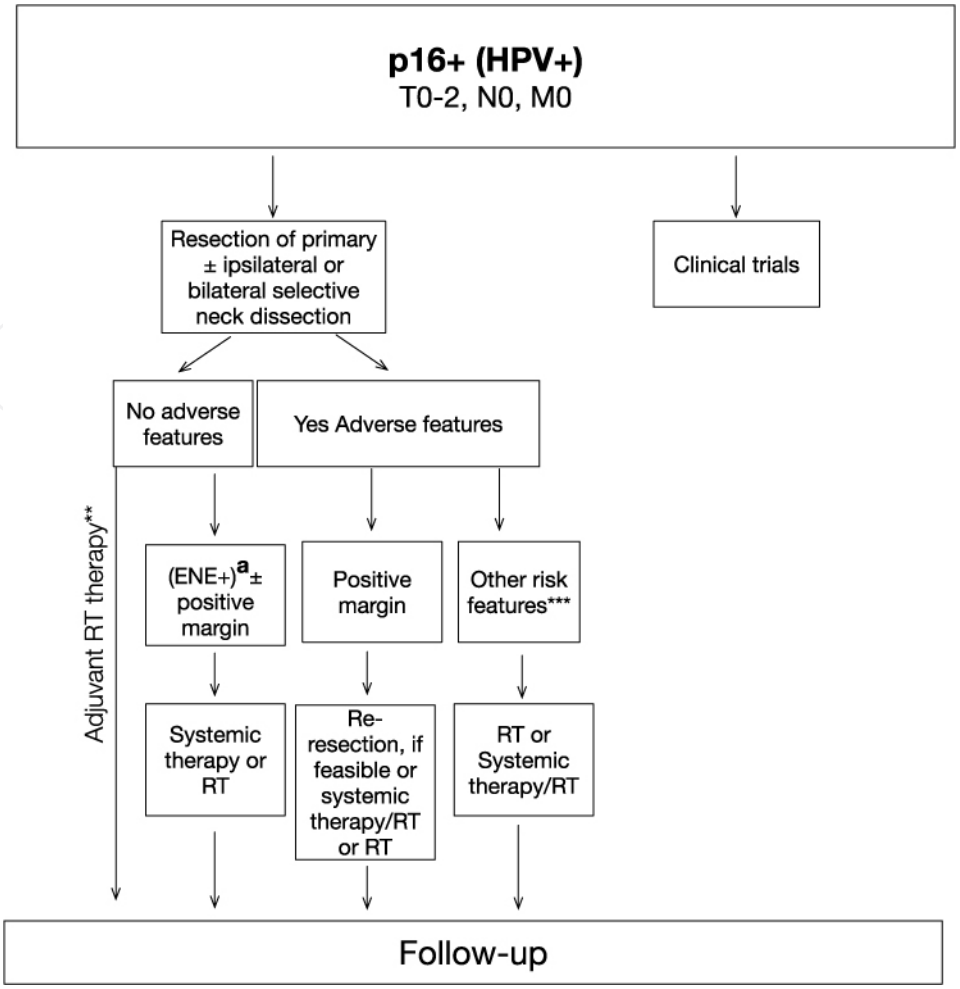

Figure 5. NCCN Guidelines Version 1.2023 for Cancer of the Oropharynx (p16 [HPV]-positive). Early Stage. *Early-stage (i.e., T1–T2N0) p16-positive oropharyngeal cancers may be treated with definitive RT or resection of the primary with neck dissection. **A prospective phase II trial of initial TORS followed by risk-adapted adjuvant treatment demonstrated a 2-year PFS rate of 96.9% for low-risk disease with TORS alone, 94.9% for intermediate-risk disease with 50 Gy adjuvant RT, 96% for intermediate-risk disease with 60 Gy adjuvant RT, and 90.7% for high-risk disease with 66 Gy adjuvant RT with concurrent weekly cisplatin. Ferris RL, Flamand Y, Weinstein GS, *et al.* Phase II randomized trial of transoral surgery and low-dose intensity modulated radiation therapy in resectable p161 locally advanced oropharynx cancer: an ECOG-ACRIN Cancer Research Group Trial (E3311). J Clin Oncol 2022;40:138–149. ***Adverse pathologic features: extranodal extension, positive margins, close margins, pT3 or pT4 primary, one positive node >3 cm or multiple positive nodes, nodal disease in levels IV or V, perineural Invasion, vascular Invasion, and lymphatic invasion (see discussion). The definition of an adverse pathologic feature in the context of HPV+ disease is an area of active research. This includes the presence and extent of extranodal extension and the number of involved nodes. [a]Lymph node extracapsular extension.

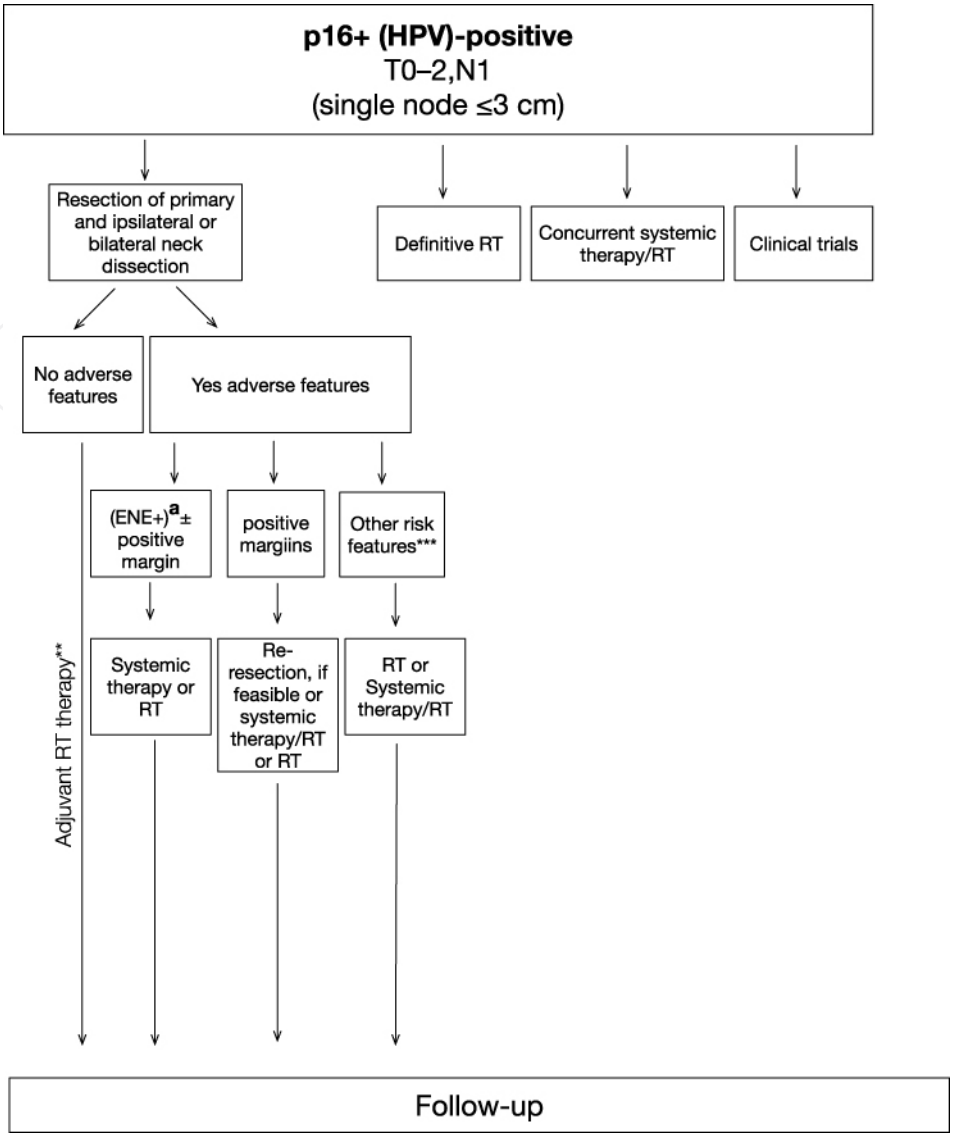

Figure 6. NCCN Guidelines Version 1.2023 for Cancer of the Oropharynx (p16 [HPV]-positive). Higher risk stage.

When ChatGPT-4 was asked to try again after delivering an incorrect answer, it frequently returned a response starting with "I apologize; however, I am restricted to the information I am trained on up to September 2021" We should recall that AJCC Stages generally represent prognoses, not specific treatment recommendations! However, as the clinical stage increases, so goes the intensity of recommended radiation [28] and or chemoradiation [22] and, at times, immunotherapy [29] is added for higher stages or resistant disease. Figures 4–8.

The authors believe that the three AIs used here (ChatGPT-4, BARD and YOU.com) should not be allowed to go unbridled in clinical medicine, and importantly these AIs are very conscious of that and are constantly reminding the users of that fact.

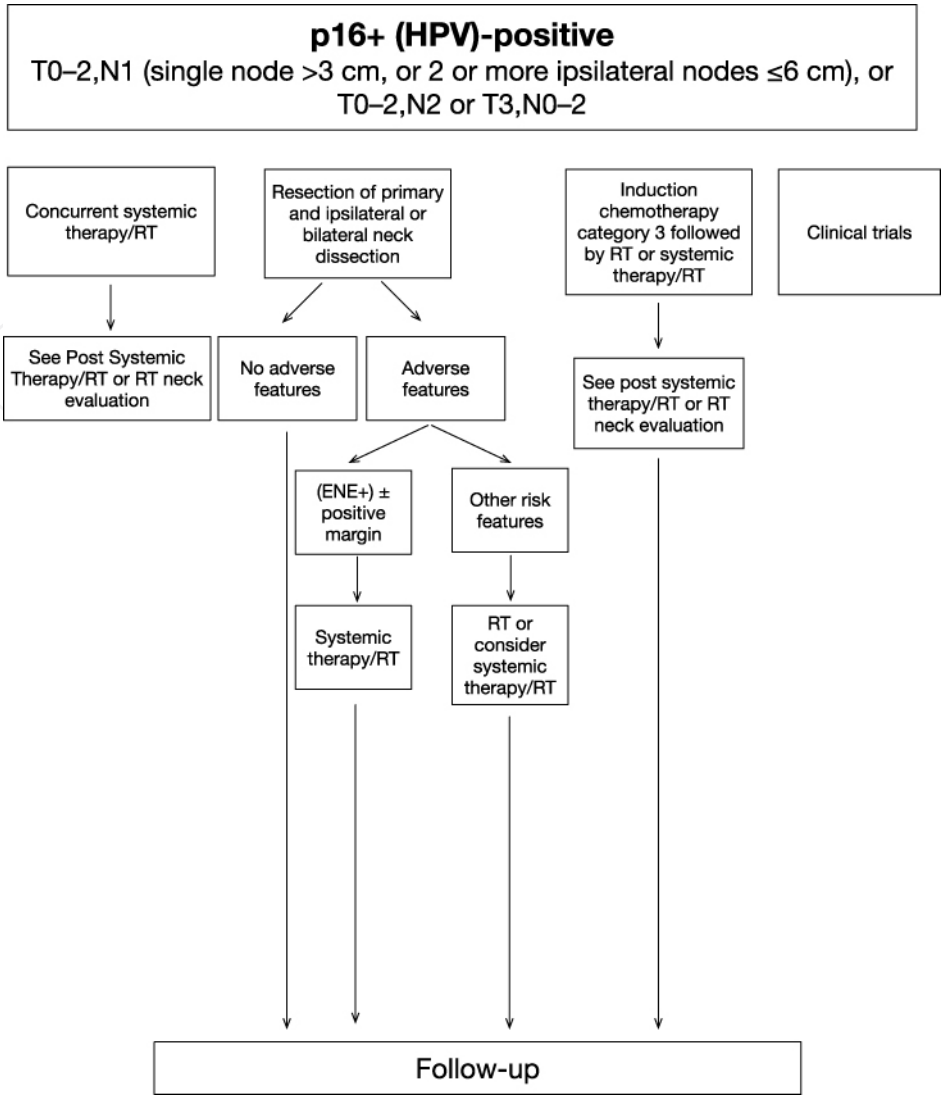

Figure 7. NCCN Guidelines Version 1.2023 for Cancer of the Oropharynx (p16 [HPV]-positive). Higher risk stage.

The choice of surgery, chemotherapy, and radiation intensity depends on accurate clinical staging, at the least. Clinical staging was a complete failure, as noted in our first attempts with ChatGPT-4. As we introduced similar prompts to the BARD and YOU.com AIs, there were similar failures despite intense refining of prompts to yield complete responses related to OPSCC staging. Prompt engineering is a process of refining interactions with AI systems, such as chatbots, to produce optimal responses. We anticipate that many physicians will not have the time to become prompt engineers; however, a useful prompt should not produce false references, a common issue with all three AIs discussed.

At this stage of AI development, inaccurate evaluations in the field of clinical medicine should not be counted on. We observed that response accuracy varied and

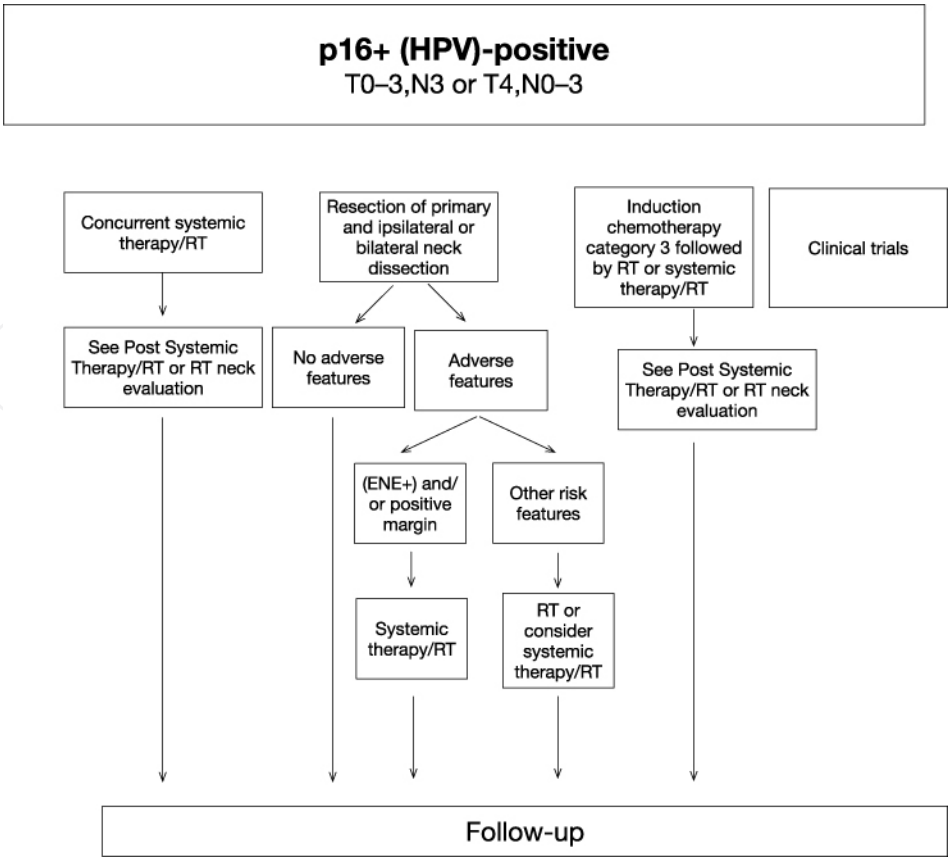

Figure 8. NCCN Guidelines Version 1.2023 for Cancer of the Oropharynx (p16 [HPV]-positive). Higher risk stage.

highly depended upon the text presented within the prompt. Medical professional users must always be prepared to check facts from fiction whenever using AI!

6.1. Other observations that are of critical importance

(A) More times than we liked, we were surprised to find that BARD or any of the other two AIs tested would suggest literature references when asked that, at first blush looked prim and proper with both DOI and PMID numbers. However, both numbers would frequently lead to entirely unrelated manuscripts or could not be found in any other available medical libraries we searched. This was a repeated problem that occurred within all three AIs tested and was very disappointing. When an unfindable (through DOI and PMID or paper title) reference offered by one AI was presented to another AI, the second AI surprisingly stated that the reference was made up and not real and did not exist. This nonexistent unfindable reference issue was universal between all AIs tested! When BARD was directly confronted with this problem through its own AI, it apologized and admitted BARD had a reference problem that they were working on fixing. The problem is

that as all three AIs tested, they usually produced the references as though they were correct, and frequently they were either incorrect or floating in a cloud somewhere.

(B) We submitted a prompt to BARD recently asking for the latest 2023 ASCO guidelines for the treatment of male breast cancer, and it came back with "...**the patient should be treated with hormonal therapy regardless of the presence or absence of hormone receptor**..." We challenged the response by asking if BARD was sure about using hormonal therapy regardless of the presence or absence of hormone therapy. BARD returned an apology revealing its mistake, indicating that hormone therapy should be restricted to hormone receptor-positive male breast cancer. If a male breast cancer patent with a hormone receptor negative tumor were treated with hormonal therapy rather than immediately initiating chemotherapy, we suspect that the patient's survival would have been compromised.

Conflict of interest

The authors declare no conflict of interest.

Appendix

We submitted an additional prompt to ChatGPT-4, and it was asked to return the TNM staging for the AJCC 7th Ed for oropharyngeal cancer as MY PROMPT = Please send me the AJCC 7th Ed of the staging manual for oropharyngeal cancer. Note below that ChatGPT-4 correctly returned the AJCC 7th Ed, which accurately did not incorporate the influence of HPV (Human Papilloma Virus (p16) positivity).

ChatGPT-4 answer:

I am an AI language model, unable to send files or attachments, but I can provide you with a summary of the AJCC 7th Edition staging system for oropharyngeal cancer.

The American Joint Committee on Cancer ***(AJCC) 7th Edition*** *divides the staging of oropharyngeal cancer into four main stages (Stage I–IV (A, B, C)) based on the tumor size (T), involvement of lymph nodes (N), and presence of distant metastasis (M). Here's a summary of the TNM classification:*

(AJCC) 7th edition

Primary Tumor (T):

- TX: Primary tumor cannot be assessed
- T0: No evidence of primary tumor
- Tis: Carcinoma in situ
- T1: Tumor ≤2 cm in greatest dimension
- T2: Tumor >2 cm but ≤4 cm in greatest dimension

- T3: Tumor >4 cm in greatest dimension
- T4a: Tumor invades the larynx, extrinsic muscle of the tongue, medial pterygoid, hard palate, or mandible.
- T4b: Tumor invades the lateral pterygoid muscle, pterygoid plates, or skull base and/or encases the internal carotid artery.

Regional Lymph Nodes (N):

- NX: Regional lymph nodes cannot be assessed.
- N0: No regional lymph node metastasis
- N1: Metastasis in a single ipsilateral lymph node, ≤3 cm in greatest dimension

N2:

- – N2a: Metastasis in a single ipsilateral lymph node, >3 cm but ≤6 cm in greatest dimension
  – N2b: Metastasis in multiple ipsilateral lymph nodes, none >6 cm in greatest dimension
  – N2c: Metastasis in bilateral or contralateral lymph nodes, none >6 cm in greatest dimension
- N3: Metastasis in a lymph node >6 cm in greatest dimension

Distant Metastasis (M):

- MX: Distant metastasis cannot be assessed.
- M0: No distant metastasis
- M1: Distant metastasis

These factors are then combined to create the overall stage groupings:

- Stage I: T1, N0, M0
- Stage II: T2, N0, M0
- Stage III: T1–T3, N1, M0 OR T3, N0, M0
- Stage IVA:
  – T4a, N0–N1, M0
  – T1–T4a, N2, M0
- Stage IVB:
  – T4b, Any N, M0
  – Any T, N3, M0
- Stage IVC: Any T, Any N, M1

Please note that this information is based on the AJCC 7th edition, and there may be updates or changes in the more recent editions. Always consult the latest edition for the most accurate information.

END

Tonsil malignancy is staged as oropharyngeal cancer according to the AJCC TNM classification of malignant tumors. The following is the **2016 8th Ed** that splits oropharyngeal cancer into (p16+) positive and (p16-) negative cancers to reflect the current understanding of the influence of HPV and (p16) status on the prognosis and management. This marks a dramatic change from previous editions and can significantly alter the final staging of the malignancy. (p16±) is considered an approximate surrogate for HPV± [30]. (**NOTE: one of every eight p16-positive OPSCCs has been found to be HPV DNA-negative. These cases have a significantly less favorable prognosis (5-year OS 46.7%), indicating the importance of performing additional HPV DNA testing for predicting prognosis and considering treatment de-intensification** [19, 31, 32]).

(AJCC) 8th edition

T Classification Oropharyngeal Cancers

- T1: Tumor 2 cm or less
- T2: Tumor more than 2 cm but less than 4 cm
- T3: Tumor greater than 4 cm or extension into the lingual surface of epiglottis
- p16 negative tumors
  - T4a: Tumor invades the larynx, deep/extrinsic muscle of the tongue, medial pterygoid, hard palate, or mandible.
  - T4b: Tumor invades lateral pterygoid muscle, pterygoid plates, lateral nasopharynx, skull base; or encases carotid artery
- p16 positive tumors
  - T4: Larynx, deep/extrinsic muscle of tongue, medial pterygoid, hard palate, mandible, lateral pterygoid muscle, pterygoid plates, lateral nasopharynx, skull base, or encases carotid artery.

N Classification p16 Negative

- N0: No regional lymph node metastasis
- N1: Single ipsilateral node less than 3 cm
- N2
  - N2a: Single ipsilateral node greater than 3 cm but less than 6 cm
  - N2b: Multiple ipsilateral nodes less than 6 cm
  - N2c: Bilateral and contralateral nodes less than 6 cm
- N3
  - N3a: Single node greater than 6 cm
  - N3b: Single or multiple nodes with extracapsular spread

N Classification p16 Positive

- N0: No regional lymph node metastasis

- N1: Unilateral nodes all less than 6 cm
- N2: Contralateral or bilateral nodes, all less than 6 cm
- N3: Metastasis greater than 6 cm

M Classification

- M0: No distant metastasis
- M1: Distant metastasis

Clinical stages for HPV-associated or (p16+) oropharyngeal Squamous Cell Cancer (SCC) of the tonsil [11].

Stage I	T0, T1 or T2	N0 or N1	M0
Stage II	T0, T1 or T2	N2	M0
Stage II	T3	N1 or N2	M0
Stage III	T0 T1 T2 T3 or T4	N3	M0
Stage III	T4	N0 N1 N2 or N3	M0
Stage IV	ANY T	Any N	M1

Additional discussion of AJCC 8th Ed for staging oropharyngeal cancers: **Emergence of a Novel Staging System for Oropharyngeal Squamous Cell Carcinoma Based on HPV Status Dec 15, 2017, Aru Panwar, MD Erik Interval, MD**

References

1 Gates B. Will ChatGPT transform healthcare? *Nat Med [Internet]*. 2023 Mar [cited 2023 May 2];**29**(3):505–506. Available from: https://www.nature.com/articles/s41591-023-02289-5.

2 Hopkins AM, Logan JM, Kichenadasse G, Sorich MJ. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm shift. *JNCI Cancer Spectr [Internet]*. 2023 Mar 1 [cited 2023 May 3];7(2):pkad010. Available from: https://academic.oup.com/jncics/article/doi/10.1093/jncics/pkad010/7049531.

3 Cabral BP, Braga LAM, Syed-Abdul S, Mota FB. Future of artificial intelligence applications in cancer care: a global cross-sectional survey of researchers. *Curr Oncol [Internet]*. 2023 Mar 16 [cited 2023 May 3];**30**(3):3432–3446. Available from: https://www.mdpi.com/1718-7729/30/3/260.

4 Bommarito M II, Katz DM. GPT takes the bar exam [Internet]. arXiv; 2022 [cited 2023 May 7]. Available from: http://arxiv.org/abs/2212.14402.

5 Konz N, Buda M, Gu H, Saha A, Yang J, Chłędowski J, et al. A competition, benchmark, code, and data for using artificial intelligence to detect lesions in digital breast tomosynthesis. *JAMA Netw Open [Internet]*. 2023 Feb 23 [cited 2023 May 6];**6**(2):e230524. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2801740.

6 Yacoub B, Varga-Szemes A, Schoepf UJ, Kabakus IM, Baruah D, Burt JR, et al. Impact of artificial intelligence assistance on chest CT interpretation times: a prospective randomized study. *Am J Roentgenol [Internet]*. 2022 Nov [cited 2023 May 6];**219**(5):743–751. Available from: https://www.ajronline.org/doi/10.2214/AJR.22.27598.

7 Ahn JS, Ebrahimian S, McDermott S, Lee S, Naccarato L, Di Capua JF, et al. Association of artificial intelligence-aided chest radiograph interpretation with reader performance and efficiency. *JAMA Netw Open [Internet]*. 2022 Aug 31 [cited 2023 May 6];**5**(8):e2229289. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2795798.

8 Kim I, Kang K, Song Y, Kim TJ. Application of artificial intelligence in pathology: trends and challenges. *Diagnostics [Internet]*. 2022 Nov 15 [cited 2023 May 6];**12**(11):2794. Available from: https://www.mdpi.com/2075-4418/12/11/2794.

9 Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primer [Internet]*. 2020 Nov 26 [cited 2023 May 4];**6**(1):92. Available from: https://www.nature.com/articles/s41572-020-00224-3.

10 Raunak V, Menezes A, Junczys-Dowmunt M. The curious case of hallucinations in neural machine translation. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies [Internet]*. Association for Computational Linguistics; 2021 [cited 2023 May 4]. p. 1172–1183. Available from: https://aclanthology.org/2021.naacl-main.92.

11 Van Gysen K, Stevens M, Guo L, Jayamanne D, Veivers D, Wignall A, et al. Validation of the 8th edition UICC/AJCC TNM staging system for HPV-associated oropharyngeal cancer patients managed with contemporary chemoradiotherapy. *BMC Cancer [Internet]*. 2019 Dec [cited 2023 Apr 24];**19**(1):674. Available from: https://bmccancer.biomedcentral.com/articles/10.1186/s12885-019-5894-8.

12 Zanoni DK, Patel SG, Shah JP. *** Changes in the 8th edition of the American Joint Committee on Cancer (AJCC) staging of head and neck cancer: rationale and implications. *Curr Oncol Rep [Internet]*. 2019 Jun [cited 2023 May 3];**21**(6):52. Available from: http://link.springer.com/10.1007/s11912-019-0799-x.

13 Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw [Internet]*. 2015 Jan [cited 2023 May 4];**61**: 85–117. Available from: http://arxiv.org/abs/1404.7828.

14 Marklund L, Holzhauser S, De Flon C, Zupancic M, Landin D, Kolev A, et al. Survival of patients with oropharyngeal squamous cell carcinomas (OPSCC) in relation to TNM 8 – Risk of incorrect downstaging of HPV-mediated non-tonsillar, non-base of tongue carcinomas. *Eur J Cancer [Internet]*. 2020 Nov [cited 2023 Apr 24];**139**: 192–200. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0959804920304470.

15 McDowell L, Casswell G, Bressel M, Gough K, Drosdowsky A, Coleman A, et al. Patient-reported quality of life and toxicity in unilateral and bilateral radiotherapy for early-stage human papillomavirus-associated tonsillar carcinoma. *Clin Transl Radiat Oncol [Internet]*. 2020 Mar [cited 2023 May 4];**21**: 85–90. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2405630820300045.

16 Brook I. Late side effects of radiation treatment for head and neck cancer. *Radiat Oncol J [Internet]*. 2020 Jun 30 [cited 2023 May 4];**38**(2):84–92. Available from: http://e-roj.org/journal/view.php?doi=10.3857/roj.2020.00213.

17 Loree J, Burke M, Muscarella J, Alam N, Romero J, Ford D, et al. Presurgical chemotherapy for HPV-positive oropharyngeal squamous cancers. 2022 [cited 2023 May 4]. e18022.

18 Embring A, Onjukka E, Mercke C, Lax I, Berglund A, Friesland S. Dose escalation of oropharyngeal cancer: long-time follow-up and side effects. *Cancers [Internet]*. 2023 Apr 30 [cited 2023 May 7];**15**(9):2580. Available from: https://www.mdpi.com/2072-6694/15/9/2580.

19 Yamashita Y, Ikegami T, Hirakawa H, Uehara T, Deng Z, Agena S, et al. Staging and prognosis of oropharyngeal carcinoma according to the 8th Edition of the American Joint Committee on Cancer Staging Manual in human papillomavirus infection. *Eur Arch Otorhinolaryngol [Internet]*. 2019 Mar [cited 2023 May 8];**276**(3):827–836. Available from: http://link.springer.com/10.1007/s00405-018-05263-x.

20 Nauta IH, Rietbergen MM, Van Bokhoven AAJD, Bloemena E, Lissenberg-Witte BI, Heideman DAM, et al. Evaluation of the eighth TNM classification on p16-positive oropharyngeal squamous cell carcinomas in the Netherlands and the importance of additional HPV DNA testing. *Ann Oncol [Internet]*. 2018 May

[cited 2023 May 8];**29**(5):1273–1279. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0923753419345338.

21 Shibata H, Saito S, Uppaluri R. Immunotherapy for head and neck cancer: a paradigm shift from induction chemotherapy to neoadjuvant immunotherapy. *Front Oncol [Internet]*. 2021 Sep 6 [cited 2023 May 3];**11**: 727433. Available from: https://www.frontiersin.org/articles/10.3389/fonc.2021.727433/full.

22 Lee J, Kim K, Kim KH, Keum KC, Kim HR, Hong MH, et al. Treatment outcomes and radiotherapy deintensification strategies in human papillomavirus-associated tonsil cancer. *Radiat Oncol [Internet]*. 2022 Dec 20 [cited 2023 May 3];**17**(1):209. Available from: https://ro-journal.biomedcentral.com/articles/10.1186/s13014-022-02177-1.

23 Yoshida EJ, Luu M, Mallen-St. Clair J, Mita AC, Scher KS, Lu DJ, et al. Stage I HPV-positive oropharyngeal cancer: Should all patients receive similar treatments? *Cancer [Internet]*. 2020 Jan [cited 2023 Apr 24];**126**(1):58–66. Available from: https://onlinelibrary.wiley.com/doi/10.1002/cncr.32501.

24 Caudell JJ, Gillison ML, Maghami E, Spencer S, Pfister DG, Adkins D, et al. NCCN guidelines® insights: head and neck cancers, version 1.2022: featured updates to the NCCN guidelines. *J Natl Compr Canc Netw [Internet]*. 2022 Mar [cited 2023 May 5];**20**(3):224–234. Available from: https://jnccn.org/view/journals/jnccn/20/3/article-p224.xml.

25 Guidelines Detail [Internet]; [cited 2023 May 6]. Available from: https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1437.

26 Thiagarajan S, Thavarool S, Kadapa N, Nandini H. Does adverse histopathological features like perineural invasion, depth of invasion, and lymphovascular invasion warrant adjuvant treatment in early oral squamous cell carcinoma? *J Head Neck Physicians Surg [Internet]*. 2017 [cited 2023 May 7];**5**(2):71. Available from: http://www.jhnps.org/text.asp?2017/5/2/71/223765.

27 Lee DJ, Kwon MJ, Nam ES, Kwon JH, Kim JH, Rho YS, et al. Histopathologic predictors of lymph node metastasis and prognosis in tonsillar squamous cell carcinoma. *Korean J Pathol [Internet]*. 2013 [cited 2023 May 4];**47**(3):203. Available from: http://www.jpatholtm.org/journal/view.php?doi=10.4132/KoreanJPathol.2013.47.3.203.

28 Kanakamedala MR, Giri SPG, Hamilton RD, Bhanat E, Vijayakumar S. Outcomes utilizing intensity-modulated radiotherapy in oropharyngeal cancers: tonsils versus base of the tongue. *Head Neck [Internet]*. 2018 May [cited 2023 May 3];**40**(5):1034–1039. Available from: https://onlinelibrary.wiley.com/doi/10.1002/hed.25077.

29 Harrington KJ, Burtness B, Greil R, Soulières D, Tahara M, De Castro G, et al. Pembrolizumab with or without chemotherapy in recurrent or metastatic head and neck squamous cell carcinoma: updated phase III KEYNOTE-048 study results. *J Clin Oncol [Internet]*. 2023 Feb 1 [cited 2023 May 3];**41**(4):790–802. Available from: https://ascopubs.org/doi/10.1200/JCO.21.02508.

30 Grønhøj Larsen C, Gyldenløve M, Jensen DH, Therkildsen MH, Kiss K, Norrild B, et al. Correlation between human papillomavirus and p16 overexpression in oropharyngeal tumors: a systematic review. *Br J Cancer [Internet]*. 2014 Mar [cited 2023 May 7];**110**(6):1587–1594. Available from: http://www.nature.com/articles/bjc201442.

31 Fauzi FH, Hamzan NI, Rahman NA, Suraiya S, Mohamad S. Detection of human papillomavirus in oropharyngeal squamous cell carcinoma. *J Zhejiang Univ-Sci B [Internet]*. 2020 Dec [cited 2023 May 9];**21**(12):961–976. Available from: http://link.springer.com/10.1631/jzus.B2000161.

32 Mensour EA, Alam S, Mawani S, Bahig H, Lang P, Nichols A, et al. What is the future of treatment de-escalation for HPV-positive oropharyngeal cancer? A review of ongoing clinical trials. *Front Oncol [Internet]*. 2022 Dec 23 [cited 2023 May 9];**12**: 1067321. Available from: https://www.frontiersin.org/articles/10.3389/fonc.2022.1067321/full.