Radiation-induced hypothyroidism after treatment of head and neck cancer

Marianne Feen Rønjom

This review has been accepted as a thesis together with three previously published papers by University of Southern Denmark 8th of May 2015 and defended on 27th of May 2015

Tutor(s): Jørgen Johansen, Carsten Brink, Laszlo Hegedüs & Jens Overgaard

Official opponents: Remco de Bree, Julie Gehl & Jens Lauritsen

Correspondence: Department of Oncology, Odense University Hospital, Sdr. Boulevard 29, 5000 Odense C, Denmark

E-mail: m_feen_r@hotmail.com

Dan Med J 2016;63(3):B5213

THE 3 ORIGINAL PAPERS ARE

1. Rønjom MF, Brink C, Bentzen SM, Hegedüs L, Overgaard J, Johansen J. Hypothyroidism after primary radiotherapy for head and neck squamous cell carcinoma: normal tissue complication probability modeling with latent time correction. Radiother Oncol. 2013 Nov;109(2):317-22.
2. Rønjom MF, Brink C, Lorenzen EL, Hegedüs L, Johansen J. Variation of normal tissue complication probability (NTCP) estimates of radiation-induced hypothyroidism in relation to changes in delineation of the thyroid gland. Acta Oncol. 2015;54(8):1188-94.
3. Rønjom MF, Brink C, Bentzen SM, Hegedüs L, Overgaard J, Petersen JB, Primdahl H, Johansen J. External validation of a normal tissue complication probability model for radiation-induced hypothyroidism in an independent cohort. Acta Oncol. 2015 Oct;54(9):1301-9.

INTRODUCTION

Head and neck cancer

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide, with an annual incidence of approximately 560,000 cases [1]. In Denmark the 2013 incidence of head and neck cancer (i.e. laryngeal, pharyngeal, oral cavity, sino-nasal and salivary gland cancer and unknown primary) was 1,287 [2]. This incidence has been increasing over the past 25 years, also when adjusted for age [3]. The 5-year age-adjusted relative survival is 64% for men and 68% for women [4]. This rate has been relatively constant for men between 1998 and 2012, but has been improving for women.

The term head and neck cancer covers a heterogeneous group of cancer diagnoses and locations, where squamous cell carcinoma of the oral cavity, pharynx and larynx is the most common histology. The environmental etiology of HNSCC is mainly related to tobacco and alcohol intake. However, the incidence of HNSCC caused by Human Papillomavirus (HPV) has increased over the last 30 years, especially for oropharyngeal cancer [5,6]. This has led to a new subgroup of patients with HNSCC, who have different clinical characteristics and molecular biology and increased radiosensitivity [7,8].

Treatment of HNSCC

HNSCC is predominantly a loco-regional disease with less than 5% of patients having distant metastases at diagnosis [2]. Around 60% of patients have locally advanced disease (i.e. stage III and VI) at diagnosis. The main treatment options are surgery and radiotherapy (RT), either alone or in combination, and in recent years with addition of chemotherapy. RT is the primary treatment for HNSCC in Denmark [9], thus 70% of patients receive primary RT and 30% receive surgery with or without post-operative RT [2]. For the majority of patients with early stage HNSCC, single modality treatment may be sufficient for cure. For patients with advanced disease, combined modality treatment including chemo-radiation [10], modification of tumor hypoxia [11] and altered fractionation [12,13] has improved outcome. Treatment with newer biological targeted agents, like epidermal growth factor receptor (EGFR) inhibitor, in combination with definitive RT is controversial [14-17].

Over the past three decades there has been a significant technological development in the delivery of RT, from two-dimensional (2D)-RT based on X-ray images to CT-based three-dimensional conformal treatment (3D-CRT). 3D-CRT uses three or more treatment fields to deliver a homogeneous dose to the target area, and subsequent implementation of multi-leaf collimators and more advanced computer algorithms, has enabled intensity modulated RT (IMRT) and later volumetric modulated arc therapy (VMAT). With these novel treatment techniques there is potential for sparing the normal tissues while increasing radiation dose to the tumor and other target areas [18-20].

The Danish Head and Neck Cancer Group (DAHANCA) has used standardized guidelines for treatment of head and neck cancer in all centers [21] since 2002, and DAHANCA trials have made a significant contribution to optimizing RT in patients with head and neck cancer, both nationally and internationally [11,12,22]. The DAHANCA database contains a variety of clinical, demographic, and treatment-related data including symptoms, etiological factors, diagnostic methods, TNM stage, primary treatment, follow-up status, disease failure, death and cause of death of patients

diagnosed with HNSCC in Denmark since 1991 (and larynx carcinoma since 1971) [2,23]. The database provides a valuable tool for quality assurance of treatment [24] and for research, including the present PhD-project [25-29].

Radiation treatment planning

Treatment is delivered as external beam radiation using linear accelerators. In definitive RT for HNSCC in this project, standard target dose is 66-68 Gy in 33-34 fractions, 5-6 fractions/week [21].Treatment planning for head and neck cancer patients in daily practice is initiated with a therapeutic CT-scan of the patient fixated in the treatment position. The tumor volume and clinically relevant target areas are defined and delineated in the treatment planning system by the oncologist and the radiologist in collaboration. The critical normal tissues, organs at risk (OARs), are defined and delineated by the oncologist. Planning target volumes are then generated and a dose calculation is performed to ensure appropriate dose distributions to cover tumor areas during radiotherapy and to keep treatment doses below tolerance levels of the normal tissues as defined by clinical guidelines. Modern treatment planning systems allow a rather precise calculation of treatment dose to small areas (submillimeter). Knowledge of tolerance levels to ionizing radiation of different organs can be used in RT dose-planning to restrict a specific dose (dose-constraint) or treated volume to an organ using either a single value, such as the mean dose to the organ, or by multiple dose-volume constraints [30].

Toxicity

The limiting factor in RT is toxicity of the normal tissues. The greatest challenge in radiotherapy is thus to optimize disease control while minimizing toxicity to the surrounding normal tissue.

Toxicity has traditionally been divided into two categories: acute reactions, defined as effects that occur within 90 days after start of RT, and late reactions which develop ≥90 days after treatment (and up to years after RT). The relevance of this separation in grading and reporting adverse effects has been challenged internationally, however, and identification of potential new temporal patterns of injury are encouraged [31].

Common acute reactions during treatment of HNSCC are pain, mucositis, dysphagia, xerostomia, loss of taste, mucosal edema and erythema, skin erythema and moist desquamation. Common late reactions include xerostomia, dysphagia, mucosal edema and skin fibrosis [32].

Such endpoints generally have been graded on four- or five-point scales [33]. Adverse event reporting in cancer treatment, and specifically radiation oncology have been the Common Terminology Criteria for Adverse Events (CTCAE) [31] and the LENT/SOMA scale [34-36]. LENT is an acronym for Late Effects Normal Tissues, while SOMA defines toxicity from Subjective, Objective, Management-related and Analytic measures (i.e. blood test, CT or the like). Graded toxicity scorings can be converted into binary data to form the basis for analyzing radiation dose-response relationships that can be used to predict toxicity in patients undergoing radiation treatment.

Normal tissue complication probability (NTCP) modeling

Since radiation of normal tissues is inevitable during external beam RT, knowledge about tolerance levels of OARs to ionizing radiation is necessary to distribute the radiation dose appropriately to avoid unacceptable toxicity. With increasing radiation dose, the frequency (i.e. incidence) and the severity (i.e. grade) of radiation effects may increase.

In radiobiological studies, the impact of radiation dose on normal tissues (and tumor tissue) is often demonstrated by dose-response (dose-incidence) curves. Dose-response curves describe the probability of a specific response (i.e. radiation effect) as a function of dose. They generally have a sigmoid (S-) shape where the risk of a given effect goes from 0% or close to at no radiation and close to 100% at high doses. Several mathematical functions have been used to model the relationship between radiation dose and normal tissue complication probability (NTCP); most frequently the Poisson, the logistic, or the probit models [37,38]. For endpoints requiring prolonged follow-up, such as late normal tissue reactions, these models may be embedded in a so-called mixture model, where the time to occurrence of the endpoint is also taken into account [39]. Dose-constraints as described above are derived from NTCP models and can be defined as dose-volume restrictions for a clinically acceptable risk of a given radiation effect [21].

In 1991, Emami et al. [40] pooled information regarding partial or total organ tolerance doses with the clinical experience and judgments of clinicians and researchers, and published tolerance doses for various organs, including the thyroid gland. The subsequent development of 3D-CT based treatment planning and the possibility of calculating precise doses to normal tissues have enabled studies of normal tissue response in relation to different dosimetric parameters. In 2010, extended analyses applying volume modeling of normal tissue reactions in clinical radiotherapy were published in the Qualitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) papers [41]. However, data regarding normal tissue reactions of the thyroid gland was omitted by the QUANTEC group.

The thyroid gland

The thyroid gland is located in the lower region of the neck, in relation to the lower part of the larynx and upper part of the trachea. The gland consists of two lobes, right and left, which are connected through the isthmus. The size of the thyroid gland in the Danish population was 15-28 ml [42], but after the introduction of salt iodization around 2000, this size seems to have decreased [43].

The follicle cells of the thyroid gland synthesize and secrete the hormones thyroxine (T4) and triiodothyronine (T3). These thyroid hormones regulate the metabolism of tissues and organs in man, and are tightly genetically regulated in healthy individuals, as shown e.g. in twin studies [44]. Thyroid hormones also influence most of the pathways of intermediary metabolism. If thyroid hormones are deficient, pathways of carbohydrate, lipid and protein metabolism are slowed, as is pathways' responsiveness to other regulatory factors. Thyroid hormones are also essential for normal body growth and the development of the central nervous system, especially in the pre- and postnatal period.

Thyroid function is regulated by a delicate negative feedback loop between mainly the pituitary and the thyroid through secretion of thyrotropin (TSH) from the anterior pituitary, and T4 and T3 from the thyroid. TSH secretion is stimulated by thyrotropin-releasing hormone (TRH) from the hypothalamus but a complicated network of hormonal and non-hormonal factors, which will not be dealt with here, influences the serum TSH level. The hypothalamic-pituitary-thyroid axis follows a diurnal circadian rhythm, with

low TSH levels in the daytime and higher levels at night [45,46] in individuals following a normal day-night sleep-awake rhythm.

Pathophysiology and mechanism of radiation-induced injury

Damage to the thyroid gland by ionizing RT is due to a variety of pathological mechanisms. These include vascular effects in the epithelium of small vessels and the development of fibrosis of capsular structures, while the damage to the follicular epithelial cells is considered less important [47,48]. Ultrasonography has shown that changes in both the blood vessels and the gland echogenicity occur during RT, and subsequent development of acute thyroiditis was correlated to vessel changes [49]. The predominant late morphological changes consist of atrophy, chronic inflammation (thyroiditis) with lymphocytic infiltration, vascular fibrosis, and focal and irregular follicular hyperplasia [47,50,51]. The mechanisms for radiation-induced hypothyroidism are largely unknown but are assumed to be related to vascular damage [48].

Hypothyroidism

Hypothyroidism (HT) is defined as decreased function of the thyroid gland and is characterized by TSH above the normal range (at Odense University Hospital the normal range is 0.3-4.0 mIU/l [52]). HT can be subclinical HT (SHT) with elevated TSH and normal T4 and T3, or overt HT with elevated TSH and T4 and/or T3 below the normal range.

Hypothyroidism, including SHT, has been linked to a number of metabolic changes such as hyperlipidemia [53], coagulopathy [54], endothelial dysfunction [55] and other cardiovascular disorders [56,57]. These changes are associated with increased morbidity such as cardiovascular diseases, lung diseases and diabetes mellitus [58]. HT has also been associated with psychiatric morbidity [59] and impaired quality of life [60,61]. Although effects on mortality have been uncertain [62], two recent Danish studies have found increased mortality in patients with overt HT compared to euthyroid controls [63,64]. While SHT has been linked to an increased risk of coronary heart disease [65,66] and mortality [67-70], the effect on overall mortality is still a matter of debate [64,66,67,69,71] and duration of thyroid dysfunction may matter [64]. For these reasons, radiation to the thyroid gland must be considered in RT of head and neck cancer.

The prevalence of HT is 1-2% [72]. In Denmark, the lifetime risk of overt HT has been reported to be 2.7% (with a 3:1 female predominance) [73]. The prevalence of SHT increases with age and in the age group 60-65 years the prevalence is 7.1% for women and 1.6% for men [74]. Substitution treatment with Levothyroxine is indicated in overt HT and some cases of SHT [75].

Development of HT is influenced by a number of factors such as gender [73], genetic predispositions [76] and environmental factors of which iodine intake and cigarette smoking are best characterized [77-79]. In a Danish study of overt HT, Carlé et al. [73] found that the dominant cause of HT was spontaneous HT (85%), mostly of autoimmune origin, while non-spontaneous HT was caused by post-partum thyroiditis (4.7%), medications such as Amiodarone (4.0%) or Lithium (1.6%), subacute thyroiditis (1.9%), congenital (1.6%), and previous treatment with radioiodine, external beam radiation and surgery (1.8%). Importantly, the prevalence of HT after external beam radiation may well be underestimated, since measurements of TSH after RT have until now not been performed routinely in Danish oncology centers.

Background for the PhD project

Radiation- induced HT (RIHT) was a recognized late effect of RT to the neck when this PhD project was initiated [47,80-95]. However, the reported incidences varied significantly as the design, endpoint, and treatment regimes in previous studies were different. Overt HT was reported in 6-20% of head and neck cancer patients [80,81,83,85,90,95] and SHT in 24-50% [82,87,88,91-93,95]. The role of chemotherapy as a risk factor of RIHT was unclear [84-86], but surgery seemed to be critical [81-83,87,89,94,95]. Table 1 shows an overview of studies.

No relevant Danish data was available on the incidence of RIHT, and TSH assessment after RT had not been part of the national routine follow-up of head and neck cancer patients. Furthermore, the tolerance dose of the thyroid gland to ionizing radiation was poorly defined. Data had been derived from only a few studies that were able to report calculated absorbed doses to the thyroid gland (see Table 1). Consequently, the radiotherapy units of Danish oncology centers did not routinely consider calculated radiation doses to the thyroid gland when optimizing radiation treatment plans.

Aims

Overall aims of the PhD project

1. To estimate the incidence of biochemical hypothyroidism in a large cohort of patients with head and neck squamous cell carcinoma (HNSCC) of the oral cavity, oropharynx, hypopharynx, and larynx treated with definitive radiotherapy and no surgery, and to assess the time-course of the development of hypothyroidism.
2. To establish a dose-response relationship for biochemical hypothyroidism by assessing the absorbed radiation doses in the thyroid gland in individual patients and thus to estimate the tolerance level of the thyroid gland to ionizing radiation.

Aims of the studies

Study I:

1. To identify risk factors for hypothyroidism after definitive radiotherapy for HNSCC.
2. To establish an NTCP dose-response relationship for biochemical hypothyroidism.
3. To provide clinical recommendations for thyroid constraints in dose planning for radiotherapy.

Study II: **Validation of input data**

1. To evaluate the variability of estimated NTCP values from the mixture NTCP model derived in study I and to assess the intra- and inter-observer variability in delineation of the thyroid gland.

Study III: **Validation of the NTCP model**

1. To investigate risk factors for hypothyroidism in a new, independent population of HNSCC patients.
2. To generate an (equivalent) NTCP model for comparison with the original data set to test the robustness of the previous NTCP model.

Table 1. Overview over studies on radiation-induced hypothyroidism (RIHT). (Studies published after the start of this PhD project have a grey background).

First author, year	No. of patients analyzed (diagnosis)	Surgery included	Endpoint	Baseline thyroid status	Thyroid RT dose calculated	Incidence/risk of HT	Risk factors of HT	Constraint or NTCP model
Hancock, 1991[84]	1,469 (HD)	No	Biochem. HT	No	No	20 year risk: 52% 26 year risk: 67%	Sex, Ch-RT, age, high radiation dose (for patients >17 years)	No
Grande, 1992 [95]	221 (HNSCC, 12 NPC)	Yes (n=106)	Biochem. HT	No	No	Crude: 41% biochemical HT (SHT 26% and overt 15%)	RT dose, combined RT and cervical surgery, time from RT and no midline shield	No
Turner, 1995 [87]	84 (H&N, 14 NPC)	Yes (laryngectomy n=14)	Biochem. HT (overt and SHT)	In 20 patients	No	Crude: 20% biochemical HT	Total laryngectomy	No
Kuten, 1996* [93]	84 (43 H&N, 41 HD)	No	Biochem. HT, (overt and SHT)	No	No	Crude: 45% biochemical HT (9/38 overt HT and 29/38 SHT)		Around 40 GY
August, 1996 [90]	35 (oral cavity SCC)	No	Overt HT	No	No	Crude: 14.3%	Radiation dose	No
Tell, 1997* [80]	264 (H&N, 11 NPC)	Yes (n=122)	SHT and overt HT	Yes	No	3-year risk: 15% overt HT, 40% SHT	The whole gland in treatment field, laryngectomy and age	No
Kumpulainen, 2000 [82]	72 (laryngeal cancer)	Yes (n=20)	Biochem. HT	No	No	Crude: 24% biochemical HT	Radiation field >7 cm, combination RT and surgery, quality of radiation	No
Sinard, 2000* [83]	198 (HNC)	Yes (6 groups, surgery ± RT)	TSH >10 mU/l	Yes	No	Crude: 15% biochemical HT		No
Colevas, 2001 [92]	118 (HNSCC)	No	Biochem. HT	No	No	Estimated 6-months: 14% 1-year: 27%	Age >60 years	No
Mercado, 2001 [88]	143 (HNSCC)	No	Biochem. HT (TSH >5.5)	Yes	No	5-year: 48% 8-year: 67%	Race (no Afro-Americans developed HT)	No
Tell, 2004* [81]	308 (HNC, 11 NPC)	Yes (n=146)	Overt HT and SHT	Yes	No	1-year risk: 7% 5-years: 20% 10-years: 27%	Increased baseline TSH, bilateral RT and surgery to the thyroid	No
Garcia-Serra, 2005 [94]	504 (H&N), 206 with follow-up TSH	Yes	Biochem. HT	No	No	5-year incidence: 42% 10 year: 74%	Stage and total laryngectomy and RT	No
Norris, 2006 [86]	390 (OPC). 169 with follow-up TSH	?	Biochem. HT	Yes§	No	Estimated 5-year incidence: 31% (whole cohort) and 54% (169 patients)	Adjuvant chemotherapy (all patients) Neckdissection (169 patients)	50 Gy
Lo Galbo, 2007 [89]	156 (laryngeal and hypopharyngeal cancer)	Yes (n=48)	Biochem. HT	No	No	Crude: 28%	Hemithyroidectomy and presence of anti-TPO and anti-Tg antibodies	No
Alterio, 2007 [96]	73 (H&N, 26 NPC)	?	Biochem. HT	No	Yes (n=57)	Crude: 26% HT	Sex (and thyroid volume from this)	No
Bhandare, 2007 [85]	312 (H&N, 116 NPC) (197 with lower neck RT)	No lower neck surgery	Blood samples and clinical data	Not in all	Yes	Estimated 5-year incidence with lower neck RT: 32% 27% central HT.	Ch-RT and race (white) (univariable)	No
Diaz, 2010 [91]	144(H&N)	No	Biochem. HT	Yes§	Yes	1-year: 23% 3-years: 53%	Age and thyroid volume	No
Wu, 2010* [97]	408 (NPC)	No	SHT and overt HT	Yes	No	5-years: 15.7% SHT, 9% overt HT, 10-years: 19% overt HT.	Age <30, 3D-CRT (overt HT). Sex, low T-stage, 3D-CRT (SHT)	No
Lin, 2011* [98]	45 (NPC)	No	Biochem. HT	Yes	Yes	Crude: 27% HT after 18 months	RT mean dose	No
Cella, 2012 [99]	61 (HD)	No	Biochem. HT	Yes	Yes (n=50)	Estimated 2 –year risk: 43.5% 5-year: 49.1%	V_{30} >62.5%	V_{30}<62.5%
Cella, 2012 [100]	Same as above, reanalysis of data						V_{30}(%) and sex, and V_{30}(cc), sex, thyroid volume	NTCP models
Boomsma, 2012* [101]	105 (H&N)	Yes (but not thyroid)	Biochem. HT	Yes	Yes	2-year incidence: 36%	Thyroid volume and mean dose	NTCP model
Bakhshandeh, 2013* [102]	65 (H&N, 11 NPC)	Yes (n=32)	Biochem. HT	Yes	Yes	1-year incidence: 45% HT		NTCP model
Rønjom, 2013 [103]	203 (HNSCC)	No	Biochem. HT	Yes	Yes	1-year incidence: 12% 5-years: 26%	Thyroid volume and mean dose	NTCP model, constraint
Lo Galbo, 2013* [50]	137 (larynx and hypopharynx)	Yes (n= 40)	Overt HT and SHT	Yes	No	Crude: 47.4% (27.7 SHT and 19.7% overt HT)	Hemithyroidectomy, laryngectomy, neckdissection and increasing age	No
Lin, 2013* [104]	65 (NPC)	No	Biochem. HT, overt, SHT and central HT	Yes	Yes	18-months incidence: 23%	D_{mean} > 50 Gy to the thyroid and pituary gland and to thyroid gland alone.	No
Lin, 2014* [105]	50 (NPC)	No	Biochem. HT, overt, SHT and central HT	Yes	No	1 year incidence: 22%		No
Murthy, 2014* [106]	89 (HNSCC)	No	Biochem. HT, overt and SHT	Yes	Yes (n=43)	Incidence: 55%, 39% SHT and 16% overt HT	Young age, primary site, node+ and D_{100} (min dose)	No
Akgun, 2014 [107]	100 (88 H&N, 38 NPC, 12 HD))	Yes (n=30)	Biochem. HT	Yes (n=52)	Yes	Crude: 52%	Thyroid volume, D_{mean} and V_{30} (univariate)	V_{30}
Kim, 2014* [108]	114 (H&N, 29 NPC)	Yes (n=33)	Biochem. HT	Yes	Yes	Crude: 46%	V_{45} >65%	V_{45}<50%
Chyan, 2014 [109]	123 (OPC)	Yes (n=27)	Biochem. HT, overt and SHT	Yes§	Yes	Crude: 61% (54% overt HT)	Thyroid volume (univariate)	V_{45}
Sachdev, 2014 [110]	75 (HNSCC) (5 NPC)	Yes (n=16)	Biochem. HT	Yes	Yes	Crude: 33%	V_{50}>60%	V_{50}<60%

HNSCC - head and neck squamous cell carcinoma, HD - Hodgkins disease, NPC - nasopharyngeal cancer, OPC - orophanrygeal cancer, biochem. HT - biochemical hypothyroidism, RT= radiotherapy, SHT = subclinical hypothyroidism, Ch-RT- chemo-radiation, 3D-CRT - three-dimensional conformal treatment

*Prospective design, if not stated, the design was retrospective

§ Baseline thyroid disorders taken into account but not TSH assessment

MATERIAL AND METHODS

This thesis is based on three studies. The first study was a retrospective cohort study from the Department of Oncology, Odense University Hospital (OUH), studying the incidence and risk factors for RIHT and development of an NTCP model for RIHT after definitive radiotherapy for HNSCC. The second study was a validation of the input data in Study I, assessing the impact of observer variability in delineation of the thyroid gland on the estimates from the NTCP model. The third study was a validation of the NTCP model from Study I on a similar cohort of HNSCC patients treated at another center, the Department of Oncology at Aarhus University Hospital (AUH).
An overview of the designs, methods, data and analyses are presented in Table 2. Further details are provided in the following sections.

Table 2. Overview of the methods used in the three studies of this thesis.

	Paper I	Paper II	Paper III
Design	Retrospective cohort study with both prospective and retrospective collection of TSH assessments	Study of intra- and inter-observer variability.	Retrospective cohort study with both prospective and retrospective collection of TSH assessments
Data source	DAHANCA database, patients treated for HNSCC at the Department of Oncology, OUH during 2002-2010. Patient charts. Laboratory database. Clinical dose planning system.	All CT scans from dose-planning RT treatment plans with thyroid gland delineations from Study I	Consecutive TSH assessments from patients seen in the out-patient clinic June to December 2012 at the Department of Oncology, AUH. Patient charts and DAHANCA database Clinical dose planning system. Blood bank at the Department of Experimental Clinical Oncology, AUH
Collected data	Patient-related Tumor-related Thyroid status (baseline and follow-up) Dose-volume parameters of the thyroid gland	$V_{thyroid}$ and D_{mean} from dose-volume histograms (DVH).	Patient-related Tumor-related Thyroid status (baseline and follow-up) Dose-volume parameters of the thyroid gland
Outcome	Incidence of HT Risk factors for HT NTCP models for HT Dose-constraints for the thyroid	Variability in delineated $V_{thyroid}$, D_{mean} and estimated risk of HT.	Risk factors for HT NTCP models for HT
Analyses	Descriptive Logistic regression Mixture modeling with Weibull distribution Bootstrap	Descriptive Bland-Altman plots Dice-Sørensen similarity index Models of spatial 3D variation	Descriptive Monte Carlo sampling Mixture modeling

Study I

Study population

Participants were patients with HNSCC in the oral cavity, oropharynx, hypopharynx or larynx treated with definitive RT (66-68 Gy in 33-34 fractions) with or without concomitant chemotherapy at the Department of Oncology, OUH, in the period 2002-2010 (cohort1). Inclusion- and exclusion criteria and patient characteristics are described in Paper I.
2002 was chosen as the start of the study period, as all head and neck cancer patients treated at OUH from 2002 received treatment according to a CT-based 3D treatment plan that had been stored after treatment.

Material

The study was designed as a retrospective cohort study with both prospective and retrospective collection of TSH assessments. HNSCC patients treated in the period 2002-2010 were identified from the DAHANCA database, see Figure 1. The clinical data from this database were crosschecked with patient charts. Missing data values were corrected, and information about thyroid disease and medication were collected.

TSH assessment

Patients were considered euthyroid if they had a normal TSH measurement within the last year before RT or in the first two weeks of RT (baseline TSH). Follow-up TSH assessments were obtained after the end of RT.
All TSH analyses were done at the Department of Biochemistry and Pharmacology (KBF) at OUH and the data were extracted from the OUH laboratory database (Netlab). All TSH assessments in the database were taken into account (i.e. from all OUH departments and from general practitioners in the county of Funen). The time period for data extraction was January 2001- March 2012.
Until 2006, serum TSH was measured using a time-resolving fluoroimmunometric assay (AutoDELFIA, Perkin Elmer, Turku, Finland), reference interval 0.30-4.0 mIU/l [52]. The intra- and inter-assay coefficients of variation (CVs) at serum TSH concentrations 0.046-17.6 mIU/l were 1.3-4.7% and 1.7-3.7%, respectively. From 2006 to September 2010, serum TSH was analyzed using a solid-phase, two-site chemiluminescent immunometric assay (third generation assay) on Immunolite 2000 equipment (Siemens, Erlangen, Germany). The intra- and inter-assay CVs at serum TSH concentrations 0.32-39 mIU/l were 3.8-5.3% and 4.5-6.4%, respectively. After September 2010, plasma TSH has been analyzed using a different third generation assay (Architect i system, Abbot, Wiesbaden, Germany). The intra- and inter-assay CVs at plasma TSH concentrations 0.088-28.4 mIU/l were 1.4-5.0% and 1.7-5.3%, respectively. To ensure compatibility between the different equipment, comparisons were done at KBF when equipment was replaced, and the same reference interval for TSH has been applied for all three methods.
Before 2008, TSH assessments of HNSCC patients were obtained at the Department of Oncology, OUH, when indicated. In 2008, a guideline for routine assessment of thyroid function in HNSCC patients was implemented, i.e. TSH and T3/T4 when indicated, before RT and once a year after RT. To include more patients from the first half of the study period, additional baseline TSH assessments were obtained from blood bank samples collected from participants in the DAHANCA 10 trial [111] (EDTA-plasma stored at – 80°C, analyzed in June 2011). Blood samples from 45 patients were analyzed, and 27 of these were included in the study. Analyses of TSH on EDTA-plasma compared to Li-heparin plasma were validated at KBF, OUH, by collecting blood samples from 20 healthy individuals (10 men and 10 women) aged 20-70 years. Analyses were done using the standard assay on Architect and showed no significant differences between the two types of plasma (p=0.423).
Patients eligible for this study who had finished scheduled follow-ups were invited for additional TSH assessment. Of the 38 patients invited, 23 (61%) accepted and gave a blood sample after informed consent.

Delineation of the thyroid gland

Treatment dose-planning was done in Pinnacle3 (Philips Healthcare, The Netherlands). The thyroid gland was not previously delineated, so the author (MFR) delineated the gland in all treatment plans. After delineation, the precise treatment dose to the gland was calculated and the corresponding dose-volume histograms (DVH) were exported for analysis. Thyroid gland volume ($V_{thyroid}$) and mean (D_{mean}), maximal (D_{max}) and minimal dose (D_{min}) as well as V_{10} (defined as the percentage volume receiving 10 Gy or more), V_{20}, etc. up to V_{70} were extracted from the DVHs.

Five of the included patients received treatment from two treatment plans. Here, the dose to the thyroid gland was evaluated in each case and if the dose per fraction was not significantly different in the two plans, the treatment plan used for the majority of fractions was chosen (n=4). Otherwise, the doses were corrected for the number of fractions given and summarized (n=1).

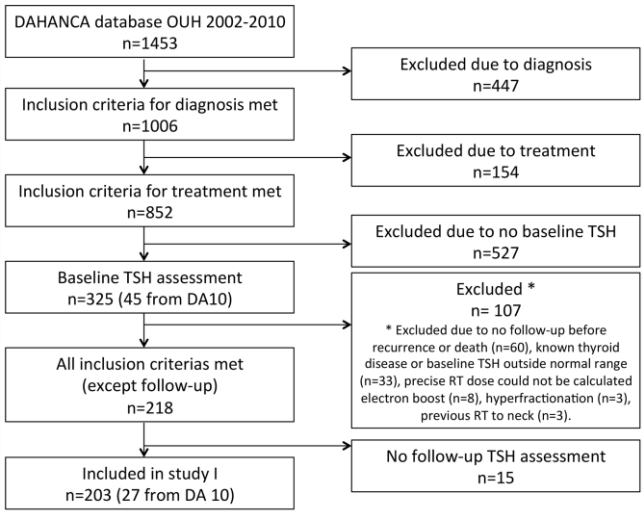

Figure 1. Inclusion of patients in Study I

Endpoint

The endpoint of the study was biochemical hypothyroidism, defined as plasma-TSH above the normal range. The normal range in the OUH laboratory is 0.30-4.0 mIU/l [52]. The first TSH assessment above the normal range was defined as an event. Follow-up was defined as time from the last day of RT to the last TSH assessment or censoring.

Statistical methods

Data were initially analyzed using standard logistic regression. Univariable analyses were performed first. Subsequently, the two most significant factors were entered into a multivariable model and the significance of the other covariates was tested by entering them into the model one by one. 95% confidence intervals (CI) were calculated using the profile likelihood method, and p-values were derived from the likelihood ratio (LR) test. The variables considered in the analyses were tested for possible interactions. Thereafter, data were analyzed using a mixture model. This was done using the logistic model and incorporating a latent time distribution. First, data were analyzed using two different time distributions, a Weibull distribution and a log-logistic distribution, representing two different classes of mathematical functions. Performance of the two models were evaluated by the models -2 log likelihood (-2LL) and residuals. There were no statistical differences between the two models, but the Weibull distribution showed slightly better -2LL and was therefore chosen for further analyses.

All parameters were optimized by maximum likelihood from the entire data set [39,112]. The mixture model was analyzed with univariable and multivariable analyses equivalently to the logistic model described above.

The significant variables from the multivariable analysis were applied to the NTCP models. The 95% CI for the NTCP models were derived by bootstrapping (10,000 samples).

The cumulative incidence of HT was calculated using the mixture model.

Study II

Study population

The thyroid gland was delineated in 246 clinically administered treatment plans in Study I by MFR, and 50 of these plans were randomly selected for this study.

Patient characteristics and treatment plans are described in Paper II.

Material

The study was designed to assess intra- and inter-observer variability. The first 50 treatment plans in which the thyroid gland was delineated, were excluded from the study to avoid a potential bias of a 'learning curve' in delineation. Fifty other treatment plans were selected from a randomly ordered list for i) blinded re-delineation of the thyroid gland by MFR (intra-observer) and ii) a corresponding delineation of the thyroid by a radiation oncologist, Jørgen Johansen (JJ) (inter-observer). Re-delineation was done at least three months after the first delineation. Guidelines were made for blinding the delineated structures in the dose plans and naming the structures, but no guideline for the delineation of the thyroid was given.

Thyroid D_{mean} (Gy) and $V_{thyroid}$ (cm^3) from the new delineations were obtained from the DVHs in the dose-planning system equivalent to Study I.

Statistical methods

Differences in intra- and inter-observer delineated $V_{thyroid}$, D_{mean} and estimated NTCP of HT were assessed by Bland-Altman plots [113]. Wilcoxon signed-rank test was used to check for systematic differences and Spearman's rank correlation was used to test for correlation between the paired differences and mean $V_{thyroid}$, D_{mean} and NTCP, respectively.

The correlation between the (random) variation and mean thyroid volume was tested by dividing the paired differences into five equally-sized groups, depending on mean thyroid size, and testing the SDs of these groups with Spearman's rank correlation.

The Sørensen-Dice similarity index (DSI) was calculated for overlap measure of delineated $V_{thyroid}$:

$$DSI\ (A,B) = \frac{2\,V_{A\cap B}}{V_A + V_B}$$

For assessment of spatial variations in delineated volume, the thyroid glands were manually divided into two lobes, and mapping of each lobe was done using the center of mass as origin and 100 measurement points in both latitudinal and longitudinal direction (10,000 surface points in total). The distance from the center to the contours was measured, including the sign (plus/minus) indicating whether the first contour was inside or outside the second contour. Based on these spatially resolved values, population averaged values (systematic deviations) as well as SD (random deviations) could be calculated and visualized as a color surface plot of the "mean thyroid gland" using the color to illustrate the local deviation.

Study III

Study population

Participants were patients treated with definitive RT at the Department of Oncology, Aarhus University Hospital (AUH). Patients were included after a follow-up TSH measurement that was taken when they attended the outpatient clinic for routine follow-up (cohort2). Inclusion and exclusion criteria are described in Papers I and III, and patient characteristics are presented in Paper III.

Material

The study was a retrospective cohort study to validate the NTCP model from Study I. The study was conducted at the Department of Oncology, AUH, in collaboration with Odense University Hospital (OUH).

TSH assessment

TSH had not previously been assessed routinely after RT at AUH, but was implemented as a routine assessment for patients in the outpatient clinic at the time of this study (partly on the basis of the findings from Study I). A baseline TSH was obtained. Due to the study design and lack of routine blood sampling, only one follow-up blood test was obtained from the AUH study group. To reduce bias due to analysis at different laboratories, all blood samples collected during the study time period were analyzed at the KBF at OUH. Blood samples were centrifuged, pipetted and frozen and sent to Odense for routine analysis. Blood samples were analyzed for TSH, TPO-antibodies and, if TSH was outside the normal range, total T3 and T4 and free T3 and T4. Baseline TSH and TPO-antibodies were analyzed on blood samples (lithium-heparin plasma stored at -80° C) from the blood bank at the Department of Experimental Clinical Oncology, AUH. Analyses were done in March 2013.

After collection of a blood sample, the patient's chart was reviewed to assess if the inclusion criteria were met, and the database from the blood bank was searched for a baseline blood sample. For the patients included in the study, all data were crosschecked with the DAHANCA database, and missing data values were corrected. The inclusion process is shown in Figure 2.

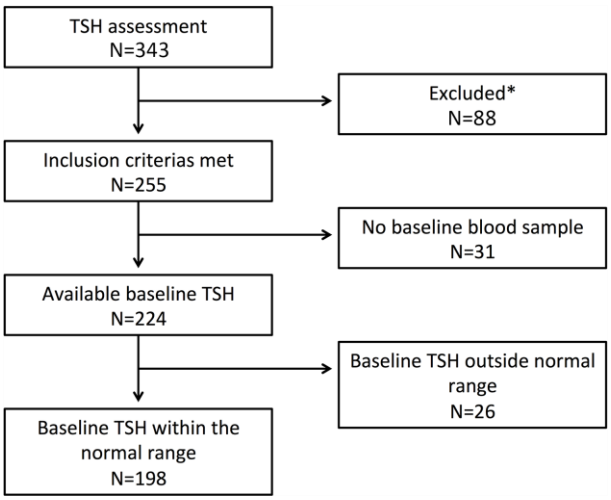

Figure 2. Inclusion of patients in Study III

Delineation of the thyroid gland and extraction of dose

The thyroid gland was delineated by the author (MFR) in all treatment plans using Eclipse (Varian, Palo Alto, USA). Thyroid volume and radiation doses were calculated. Afterward, the treatment plans were exported as DICOM RT data to the DICOM Collaboration system ("the CIRRO-server") at Odense University Hospital [114]. Thyroid D_{mean} and $V_{thyroid}$ were extracted from the recalculated DVHs in the database. The database has a built-in data validation and verification function, and the dose and volume were also double-checked with the values extracted from Eclipse after delineation.

In cohort2, 32 patients received treatment from two treatment plans or more. It was therefore necessary to consider a uniform method to calculate D_{mean} to the thyroid gland in these treatment plans. Eleven patients were treated using two or more treatment plans from one CT scan, while 21 patients were treated using two or more treatment plans from two or more CT scans. The $V_{thyroid}$ of the first clinically applied treatment plan was chosen for further analysis of volume. The different treatment plans were exported to the CIRRO server with the total dose corrected for the number of fractions given from the corresponding dose plan. D_{mean} values from multiple dose plans in an individual patient were extracted from the CIRRO server and the doses were then summarized.

Endpoint

The endpoint was biochemical hypothyroidism and follow-up was defined as the time from the last day of RT to the TSH assessment (as in Study I).

Statistical methods

The study was a validation study of the mixture NTCP model developed in Study I. Data from cohort2 were thus analyzed in the same way as in Study I, and the results were compared with cohort1.

Due to only one TSH assessment in cohort2 compared to multiple blood samples in cohort1, sampling bias was considered.

To reduce this potential bias in the comparison of the two cohorts, a new cohort (cohort1S) was generated from cohort1 by taking only one follow-up TSH-assessment into account. This was done by 10,000 Monte Carlo samplings. Analyses on cohort1S were done on each individual sample and the reported values and confidence intervals were obtained from the distribution of values from all samples (median and standard deviation).

Latency could not be estimated from cohort2, therefore, the time factor (Weibull distribution) from cohort1 was fixed and applied to both cohort1S and cohort2.

More detailed descriptions of the statistical methods are given in Papers I and III. Analyses were done using STATA 11 and 13 and MATLAB version 2012b. All statistical tests were two-sided and a p-value <0.05 was considered statistically significant.

Ethics

The studies were approved by the Regional Ethics Committees for Southern Denmark (S-20110027) and the Danish Data Protection Agency (J.nr. 2011-41-575 (Study I) and J.nr. 2012-41-0899 (Study III)). All patients invited for additional TSH measurement signed an informed consent form before the blood sample was collected.

RESULTS

The results of the three studies are summarized here. More detailed results, tables and figures are provided in the three papers (see appendix).

Hypothyroidism after primary radiotherapy for head and neck squamous cell carcinoma: normal tissue probability modeling with latent time correction (Paper I)

Two hundred and three patients were included in cohort1. The cohort was representative for all HNSCC patients treated with definitive RT at our institution in that time period (data not shown). A single follow-up TSH measurement was obtained in 96 patients and two or more (range 2-11) measurements were obtained in 112 patients. HT occurred in 35 (17%); 12 women (26%) and 23 men (15%). In 31 of these patients, free T4 or total T4

were available, and on the basis of this, five were diagnosed with overt HT and 26 with SHT.
The ultimate incidence of HT estimated from the mixture model was 26%, which corresponded to the 5-year risk. Figure 3 demonstrates that HT is most likely to develop within the first two to three years after RT.
Both logistic and mixture multivariable analysis found that higher D_{mean} and a small $V_{thyroid}$ were significant risk factors for HT. Radiation technique (concomitant boost technique in the initial treatment period vs. integrated boost (IMRT) in the last period) was a significant factor in logistic analysis, but not when taking latency into account. None of the other factors were significant in multivariable analysis, including concomitant weekly Cisplatin.
Two NTCP models were developed that included the two significant factors from the multivariable analysis, D_{mean} and $V_{thyroid}$ (Figure 4). NTCP values were calculated from the following equations:
NTCP= (1+exp(-s))-1

Logistic model:
$S=-2.019 + (0.0821 * D_{mean}) - (0.189 * V_{thyroid})$

Mixture model:
$S= -1.235 + (0.1162 * D_{mean}) - (0.2873 * V_{thyroid})$

On the basis of the mixture NTCP model, we proposed dose constraints for the thyroid gland in RT treatment planning of 26 Gy, 38 Gy, 48 Gy and 61 Gy (D_{mean}) to keep the 5-year risk of RIHT below 25%, for thyroid volumes of 10, 15, 20 and 25 cm^3, respectively.
Two methods for analyzing clinical data after RT of HNSCC patients were demonstrated, a logistic model and a mixture model that took latency into account, and these results showed the importance of taking latency into account in NTCP modeling.

Conclusion:
Biochemical hypothyroidism is a frequent late effect after radiotherapy, with an estimated 5-year incidence of 26% in a HNSCC cohort. The risk of RIHT is significantly associated with small pretreatment thyroid volumes and high thyroid mean dose. Comparison of the logistic and mixture models demonstrated the importance of latent-time correction in NTCP modeling. Individual dose constraints based on thyroid volume should be considered in radiation treatment planning.

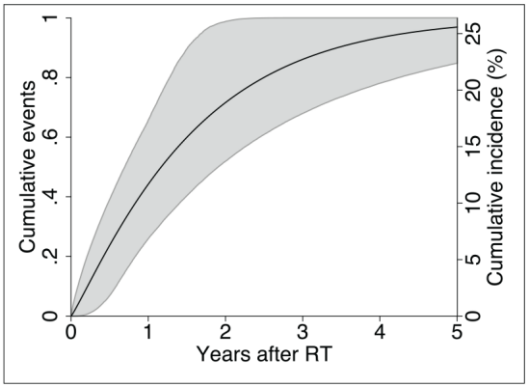

Figure 3. Cumulative events (with 95% CI) are shown on the left Y-axis, indicating the time to development of HT if HT occurs. Cumulative incidence (without 95% CI) of HT in the study population is shown on the right Y-axis.

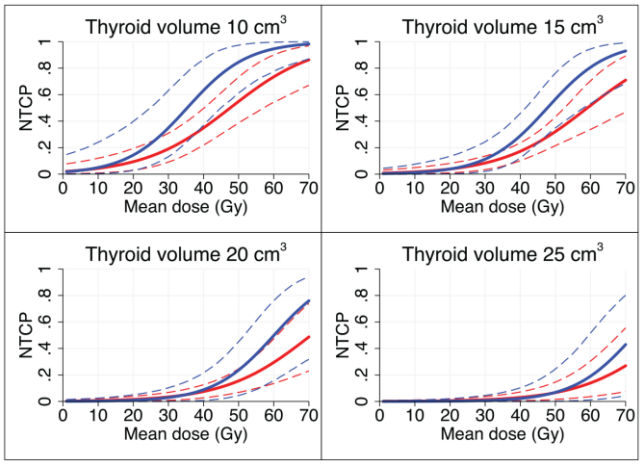

Figure 4. NTCP models at thyroid volumes of 10, 15, 20 and 25 cm^3, for the mixture model with 95% CI (blue lines) and the logistic model with 95% CI (red lines).

Variation of normal tissue complication probability (NTCP) estimates of radiation-induced hypothyroidism in relation to changes in delineation of the thyroid gland (Paper II)
This study assessed the impact of intra- and inter-observer variability in delineation of the thyroid gland on the risk estimates from the derived mixture NTCP model in Study I. Small mean (systematic) differences were found in intra- and inter-observer variability in $V_{thyroid}$, 0.4 cm^3 (SD±1.6) and 0.3 cm^3 (SD±1.8), and in D_{mean}, -0.5 Gy (SD±1.0) and 0.0 Gy (SD±1.3), respectively (Fig. 5). The corresponding mean differences of the estimated NTCP values for RIHT due to intra- and inter- observer variations were small at -0.4% (SD±6.0) and -0.7% (SD±4.8), respectively. However, as demonstrated in Fig. 5 E and F and by the relatively large SD on the estimated risk of RIHT, there may be large differences in estimated risk for the individual patient due to variability between delineations.
Mean Sørensen-Dice similarity index was 0.88 (SD± 0.03) for intra-observer delineations and 0.85 (SD ±0.04) for inter-observer delineations.
The spatial differences in delineated thyroid gland volume showed that the most pronounced mean (systematic) variability was around the middle of the thyroid gland. For the majority of the surfaces, a variability of ±0.02 cm was found in both intra- and inter-observer plots. The highest local random variations were in the caudal and medial parts of the lobes, with a local SD of ± 0.1-0.2 cm. For the majority of the surfaces, random variation with local SD of ± 0.06-0.1 cm was found. This demonstrates that the local variation in delineations is generally small.

Conclusion:
The variation in predicted risk of RIHT was small for the study population as a whole. The derived NTCP model was shown to be robust towards observer variation in delineation of the thyroid gland. However, for the individual patient, there might be pronounced differences in estimated risk due to this variation, as reflected by the relatively large SDs of the risk estimates.

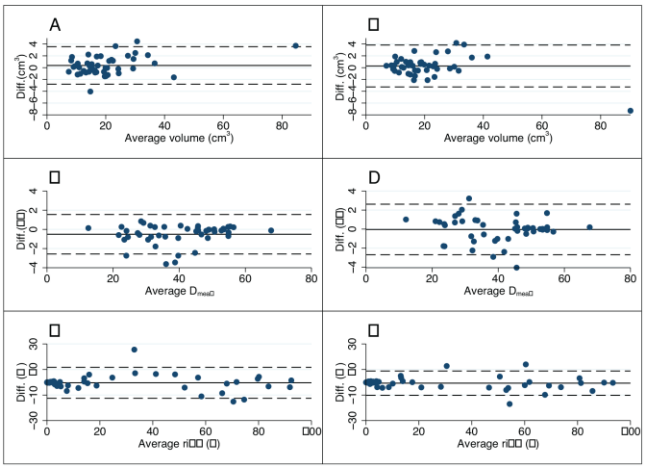

Figure 5. Bland-Altman plots demonstrating intra-observer (left column: A, C , E) and inter-observer (right column B, D, F) variation in delineated volume (A, B), calculated mean dose (C, D) and variations in estimated risk of radiation-induced HT (E, F).

Validation of a normal tissue complication (NTCP) model for radiation-induced hypothyroidism in two independent cohorts (Paper III)

A cohort of 198 patients (cohort2) was obtained after a single assessment of TSH after RT. There were no statistical differences between cohort1 and cohort2 on most patient characteristics, however, D_{mean} and baseline TSH were significantly lower in cohort2.

Analyses using the mixture model with a fixed time factor were performed on cohort1, cohort2 and the sampled cohort1S. High D_{mean} and a small $V_{thyroid}$ were found to be significant risk factors in all three cohorts. Baseline TSH was significant in cohort2; however, this was not included in the NTCP model, due to the decision to develop a model equivalent to the NTCP model from cohort1 (NTCP1).

Direct comparison of NTCP1 and the NTCP model (NTCP2) from cohort2, was not possible due to a discrepancy in the number of TSH samples. This was corrected for by developing a further NTCP model from cohort1 (NTCP1S) with just a single sample from each patient. When the number of TSH assessments was taken into account, the estimated risks were similar. Figure 6 shows dose-response curves for the three NTCP models. Differences in estimated risk between NTCP1 and NTCP2, and NTCP1S and NTCP2 are shown in Figure 7.

The NTCP model previously developed from cohort1 was:

NTCP = (1+exp (-S))-1 , where:

$S=-1.235 + (0.1162 * D_{mean}) – (0.2873 * V_{thyroid})$

The NTCP models for cohort1S and cohort2 with fixed latent time distribution resulted in the following values:

NTCP1S: $S= -4.9177 + (0.18 * D_{mean}) – (0.30 * V_{thyroid})$

NTCP2: $S= -4.7071 + (0.17 * D_{mean}) – (0.28 * V_{thyroid})$

The clinical recommendations and the latency function from study I were not validated in this study due to the differences between the cohorts.

Conclusion:

As in study I, this validation cohort showed that, a small pre-treatment thyroid volume and high mean dose were significant risk factors for RIHT. Direct comparison between the previously developed NTCP1 model and the NTCP2 model from the validation cohort was not possible. When taking the number of TSH assessments was taken into account, however, the two new models NTCP1S and NTCP2 were similar.

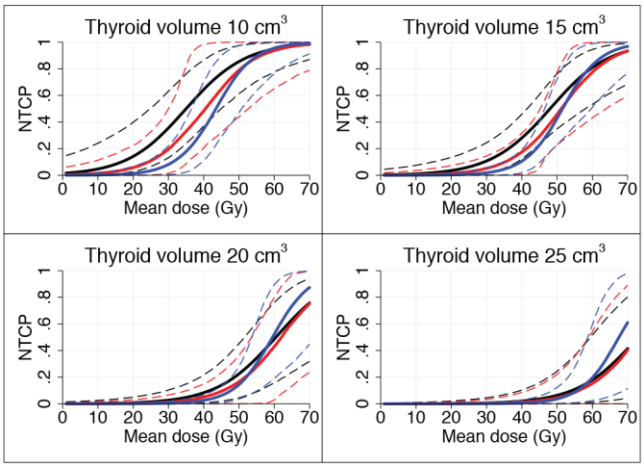

Figure 6. Dose-response curves for the three NTCP models for selected thyroid volumes: NTCP1 (black), NTCP1S (red) and NTCP2 (blue).

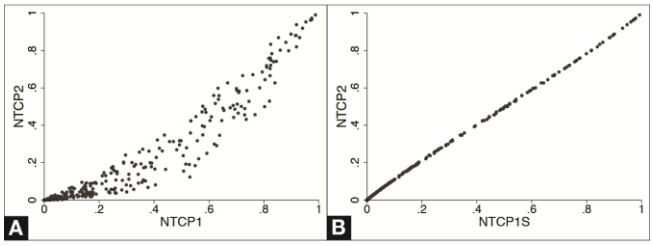

Figure 7. A) The risk of developing RIHT, estimated from NTCP1 (multiple blood samples) and NTCP 2 (single blood samples). B) The risk of RIHT, estimated from NTCP 1S and NTCP2 (both one blood sample).

DISCUSSION

The three studies in this thesis address risk predictions of radiation-induced hypothyroidism (RIHT) after treatment for head and neck squamous cell carcinoma, as well as validation issues in relation to a normal tissue complication probability (NTCP) model for RIHT.

Incidence of RIHT

Study I estimated the 5-year actuarial probability and the ultimate probability of RIHT in cohort1 to be 26% [95% CI: 18-36%]. Most of the patients who eventually developed RIHT did so within the first two to three years after RT. A peak in development of RIHT 1.5-3 years after RT is also reported in other studies [81,83,88,91,92,109]. Some have described late occurrences up to 10-27 years after RT [81,84,85,94]. However, it should be noted that these studies do not consider other causes of HT.

The actuarial probability in Study I is in accordance with findings from a number of previous studies [81,85,97], but was somewhat lower than in most studies of RIHT [88,91,94,99,101,102,107]. This lower incidence may reflect the exclusion of surgically treated patients and lower RT treatment doses, as described below. The high variability in reported incidence reflects the heterogeneity in study populations in the literature, which should be taken into account when comparing the studies. Surgery can significant-

ly impact the risk of developing RIHT [89,115] and the incidence of RIHT is likely to be higher if surgical patients are included.
The impact of a high radiation dose to the thyroid gland on the risk of RIHT is well established. This was demonstrated in Study I and III and is supported by many other studies [101,104,107-109,115]. The actual doses absorbed by the thyroid gland should thus be considered when comparing incidences of RIHT and also during treatment planning to avoid RIHT.
During the last 10-15 years, the addition of chemotherapy has become standard in curative treatment of HNSCC. However, in contrast to surgery, concomitant chemotherapy (i.e. platinum and 5-FU) of HNSCC does not seem to increase the risk of RIHT [101,102,115], as supported by our findings. Nor did we find gender to be a risk factor, in contrast to other studies [84,96,97]. Pretreatment $V_{thyroid}$ on the other hand, was found to significantly affect the risk of RIHT in Study I and III. This finding has also recently been noticed in other clinical studies [91,100,101,109]. One explanation for this is the new ability to consider and determine the thyroid volumes in CT-based dose planning systems. Alterio et al. [96] suggested that relatively higher levels of TSH were due to the smaller glands in women. This is supported by the findings in Study I, where, women had significantly smaller $V_{thyroid}$ than men, but in subgroup analyses $V_{thyroid}$ retained its significance in both men and women.

Do we need an NTCP model for RIHT?

As the cancer-specific survival for locally advanced HNSCC has doubled in Denmark since the first DAHANCA studies in the 1970s [116], the prevalence of HNSCC has increased as well. Due to the possible long-term consequences of hypothyroidism, dose constraints for the thyroid gland are needed in treatment planning for radiotherapy.

The clinically acceptable risk of RIHT is a matter of debate. However, target coverage of tumor volumes should never be compromised to reduce the dose to the thyroid gland. We propose to keep the risk of RIHT below 25% with respect to individual thyroid volumes, i.e. keeping thyroid D_{mean} below 26 Gy, 38 Gy, 48 Gy and 61 Gy for thyroid volumes of 10, 15, 20 and 25 cm^3, respectively. The recommendation of taking both D_{mean} and $V_{thyroid}$ into account is supported by Boomsma et al [101]. Others have recommended the use of threshold values for constraints, using different percentage values of V_{30} [99,100,107], V_{45} [108,109] and V_{50} [110]. However, due to the high correlation between these dose-volume parameters, we found that D_{mean} was a relevant descriptor of the dose distribution within the thyroid gland (discussed in the section “Statistical considerations”).
Knowledge about the dose-response relationship in an OAR enables us to balance the radiation dose to that organ against the dose to the surrounding normal tissues. Also of importance is that if the dose is below the tolerance level of an organ, the effect of further dose reductions may be negligible. Defining constraints for the thyroid gland allows for treatment planning optimization of thyroid doses in relation to other OARs in the neck area, including the brain [117], the pituitary [118], the parotid [119] and submandibular glands [120], swallowing organs [28,121] and the larynx [122]. Median D_{mean} was significantly lower in cohort2 than cohort1, which may be ascribed to a dose-planning strategy that had been implemented at AUH to reduce treatment dose to the larynx (to reduce the risk of radiation-induced dysphagia), and led to a lower dose to the adjacent thyroid gland. This is in accordance with two other studies that have found that doses to the thyroid gland were higher in IMRT plans than 3D-CRT, when the thyroid gland was not included as a constraint during the dose optimization process [91,106]. However, thyroid doses could be reduced and target coverage maintained using a thyroid dose constraint [91].

Uncertainties of NTCP models

Study II showed that, for the study population as a whole, the NTCP estimates of our mixture model were robust against observer variability in delineation of the thyroid gland and the corresponding variability in D_{mean}. The low variability between thyroid gland delineations may due to the thyroid gland being relatively well-defined on CT scans because of its well vascularized tissue [123] and high content of iodine [124,125]. For some individual patients, however, there were considerable differences in estimated risk of RIHT (up to 26%), especially when the estimated risk of RIHT was on the steep part of the dose-response curve. Hence, precise delineation of the thyroid gland is still of importance for consistency of NTCP estimates.
Figures 4 and 6, as well as Table 3 in Paper I, show the 95% CI of our estimated dose-response curves and suggested thyroid constraints. These demonstrate that it is important to consider not only the numerical value, but also the statistical uncertainty of an NTCP estimate or constraint in treatment planning. Furthermore, it must be kept in mind that a population-based NTCP model is only a prediction based on a standard population, thus there can be large uncertainties in the predicted probability for an individual patient, who typically deviates from the average population. Marks et al. [126] used the term “host factors” to describe specific biological features that might affect radiation sensitivity in the individual. In the current context, these might be the presence of thyroid autoantibodies or thyroid nodularity before RT. The presence of autoantibodies and their impact on the risk of RIHT is not fully elucidated [50,89,105] and should be considered in future NTCP modeling studies, or at least when assessing the individual patient risk of developing RIHT.
Despite all the uncertainties related to NTCP models, they will often be our best recommendation for considering effects on normal tissues and are therefore valuable tools for optimization of dose planning.

Choice of endpoint

Biochemical hypothyroidism was chosen as the endpoint for statistical analyses and NTCP modeling. This is the usual endpoint in RIHT studies, in both the clinical [50,82,84,86,87,91-96,109] and radiobiological [101,102] settings and corresponds to a grade 1 or more toxicity on the CTCAE classification [127] and the analytical SOMA/LENT classification [128]. A single TSH assessment above the normal range was classified as an event, as also used in epidemiological studies. So far, only one group studying RIHT has used two elevated TSH assessments as endpoint [99,100]. The choice of endpoint reflects the choice of a pragmatic, patient-centered approach to the development of NTCP models rather than a mechanistic and biology-centered approach [36]. A TSH level >4.0 (irrespective of the baseline TSH value) leads typically to clinical action such as further blood tests, additional control or follow-ups, or possibly Levothyroxine treatment. Other endpoints could also have been chosen, and a biological effect measure was in fact evaluated.
As stated in Paper I, we also analyzed the data using the change in TSH (ΔTSH >2.7 mIU/l) from baseline to follow-up as endpoint. This threshold was chosen to produce the same number of patients developing RIHT as with the clinical endpoint of TSH>4.0.

From a biological point of view, the endpoint expressing a certain change in TSH might be the best predictor of biological changes in the thyroid gland after RT. The analyses with either TSH >4.0 or a ΔTSH >2.7 as endpoint resulted in very similar dose-response curves, as most patients who experienced an event were classified as experiencing an event using either endpoint.

Uncertainties in TSH assessments

TSH assessments can vary considerably. This was not taken into account either in our studies, or in any other studies of RIHT. Uncertainties in TSH assessment can be divided into pre-analytical, analytical and biological variation [129-131]. Pre-analytical variability relates to the blood sampling procedure as well as handling and storage of blood samples. The blood samples in our studies were taken in a routine setting. Analytical variability relates to the uncertainties of the assays used for TSH analysis. Biological variation can be divided into intra- and inter-individual variations. Intra-individual variation is due to the circadian variation in serum TSH, with the lowest levels occurring in daytime and a nocturnal rise of more than 100% that peaks after midnight [132]. A decrease of up to 50% may be observed from 8:00 to 9:30 in the morning [133]. Moreover, the degree of glycosylation of TSH, non-thyroidal illness, weight change, pregnancy [134] and different medications [134,135] may affect individual TSH levels. There is substantial inter-individual variation in TSH assessments [136], demonstrated by the finding that individual 95% CI are approximately 50% of the reference range for TSH [137]. The relatively narrow individual reference range of TSH in the wide population-based reference range supports the use of ΔTSH as the endpoint for NTCP modeling of RIHT. However, as stated above, the 'patient-centered' endpoint (TSH>4.0) demonstrated similar dose-response relationships to those obtained from ΔTSH analysis.

Validity of the NTCP model

In the QUANTEC report, Bentzen et al. [41] introduced three concepts for validation of NTCP models: face validity, internal validity and external validity. Face and internal validity can be related to the mixture model (NTCP1) developed in Study I and external validity to Study III.

Face validity

Face validity relates to whether the findings of a study are reasonable. As expected in study I, an increasing radiation dose to the thyroid gland increased the risk of hypothyroidism. The dose-response relationship for development of RIHT is supported by most studies [80,82,84,95,98,99,108,109] and D_{mean} has generally been used as dose parameter in corresponding NTCP models for RIHT [101,102].

The thyroid gland is considered to be a "parallel organ" [138] with the follicle cells as functional subunits that are independent of their neighbor cells. In this context, D_{mean} is considered an appropriate dose parameter for predicting normal tissue tolerance levels, equivalent to radiobiological investigations of other parallel organs like the parotid glands [119], lung [139], and liver [140]. $V_{thyroid}$ was not expected to predict the risk of RIHT. However, as mentioned before, the importance of $V_{thyroid}$ was subsequently supported by other studies of RIHT [91,100,101].

The study findings appeared intuitively correct and the dose-response curves "made sense" in that the risk of RIHT was zero (or close to) at dose zero and the confidence intervals were reasonable in relation to our sample size and number of events. The NTCP1 model was therefore considered as having face validity.

Internal validity

Internal validity is a consideration of how well the model describes the data. This is typically assessed by plotting the observed results against those predicted by the model. However, the mixture model predicts the risk of RIHT to infinite time, while our observation period varied in length from months up to nine years. The predicted risk by the mixture model was thus expected to be higher than the observed risk. Internal validity was therefore explored in an additional way, by plotting the dose-response curves from our models in conjunction with the actual patients being either responders (equal to 1) or non-responders (equal to 0). This plot demonstrated that there were no events when D_{mean} was <25 Gy and the number of events in relation to non-events increased with increasing dose on the correspondingly rising dose-response curve.

The data seemed reasonable for both the logistic curves and the curves of the mixture model. Furthermore, when the likelihood ratio test was used to test the statistical significance of adding other parameters to the model, no other factors than D_{mean} and $V_{thyroid}$ were statistically significant for the final models. It is thus reasonable to conclude that the mixture model has internal validity in this study.

External validity

External validity relates to how well our mixture model describes other datasets. During the planning of Study III, the meaning of validating a model was considered whether the parameters of the model or the full NTCP model should be validated. The parameters of two models might be different, but if they mutually balance each other in the full model so that the risk estimates of the two models are similar, the overall clinical impact of the two models are identical which is a way to seek model validation. Validation of the full model was chosen because it was considered that risk estimates and the corresponding tolerance levels were of interest and relevance in daily clinical radiotherapy practice. The aim was then to develop an NTCP model for use in treatment planning, hence, this approach to validation of the model. Data in Study III were thus analyzed in the same way as the mixture model in Study I.

NTCP1 was not directly comparable with the NTCP model developed in the independent cohort (NTCP2), despite recruitment of another large study population. It became evident that the number of TSH assessments and the definition of endpoint had a significant impact on risk estimates.

Due to the variability in TSH assessments, the probability of experiencing an event increases with repeated assessments, when only one assessment of TSH>4.0 defines an event.

Accounting for latency by the mixture model, i.e. considering events occurring at various time-points after treatment, would allow the two study cohorts to be compared despite their differing observation times. But the latency function could not be validated in cohort2 due to only one TSH assessment. A sampling of an identical number of measurements in cohort1 allowed a comparison between the NTCP1S and NTCP2 models, and they (and the corresponding model parameters) were very similar.

The aim of validating the NTCP model was not fully achieved. However, D_{mean} and $V_{thyroid}$ were also significant risk factors in cohort2, and the NTCP models became very similar when the number of TSH assessments was taken into account. It is conclud-

ed that the NTCP model has external validity and at least, the model cannot be rejected.

Methodological and statistical considerations

Study I

A retrospective design was chosen for this study, both to obtain a study cohort of at least 150 patients and to obtain follow-up over several years for at least a proportion of the included patients. This design gave two main limitations to take into account when choosing the statistical methods for data analysis:

1. The variable number and timing of TSH assessments after radiotherapy.
2. The considerable variation in follow-up time (median 25 months, range 2-94 months), and thus the probability that some patients would develop hypothyroidism after the last TSH assessment (time censoring).

Analysis was first done using a logistic regression model. Due to the above limitations, the standard logistic regression might underestimate the risk of HT so the data were analyzed using a mixture model taking latency into account [39,112]. The accumulated Weibull distributions used for analyses represent flexible time distributions, and the time probability distribution is modeled based on the observed data. Paper I shows a comparison of data analyses using standard logistic regression and the mixture model, shown in Figure 4 [103]. The mixture model is our preferred model, and was used for modeling in Study III.

A review of the DVH's for the thyroid showed that D_{mean} was a simple descriptor for dose-distribution in the studied glands. We did not find any other parameter (i.e. V_{20}, V_{30}, etc.) to be a better predictor of the risk of RIHT. For example, if high doses were more harmful that could be demonstrated by D_{mean} (unpublished data). D_{mean} was highly correlated with the other dose-volume parameters.

NTCP modeling was done using physical doses and not fraction-size corrected doses, due to the lack of quantitative estimates of the fraction sensitivity (α/β-ratio) of the thyroid gland. An attempt was made to estimate the α/β-ratio for the thyroid gland from the dose distributions found in this study. However, this was not possible due to the small variations in dose distribution in the glands receiving almost the same D_{mean}. Consequently, the DVH's could not be reduced to equivalent uniform doses used for estimation of more advanced models like a Lyman-Kutcher-Burman model (e.g. to evaluate the adverse effects of a high dose to a small part of the gland, compared to a low dose to the whole gland). By using physical doses, the risk estimates are directly applicable in the clinical setting.

The 95% CIs for the NTCP models were derived by bootstrap (Monte Carlo simulation). Bootstrap is a non-parametric method that allows estimation of the size of the model's uncertainties. This is done by drawing new samples with replacement from the original data, under the assumption that the study sample is representative for the entire population [141,142].

The **strength** of the study was inclusion of the largest cohort used for NTCP modeling of RIHT, comprising only patients treated with definitive RT and without surgery. Additionally, the cohort was representative for the patient population treated at our department in that period, with three-quarters of the cohort being men. The exact radiation doses to the thyroid gland were calculated from the clinical treatment planning system. Delineation of the thyroid gland by one observer is thought to be a strength in modeling studies, but this can be discussed in the light of our findings in Study II.

The **limitations** of the study relate to its retrospective design, the considerable variation in number of TSH assessments and the length of follow-up. Using the hospital database is a possible source of selection bias, There may have been over-selection of patients with symptoms of hypothyroidism (who had TSH assessed) resulting in a potential overestimation of RIHT, especially before 2008 when routine assessment of TSH after radiotherapy was established. However, the criteria of a normal baseline TSH and inclusion of patients from the DAHANCA 10 study should partly compensate for this effect.

Study II

This study was designed to investigate intra- and inter-observer variability in delineation of the thyroid gland, and subsequent variability in calculated dose to the thyroid gland and estimated risk of RIHT by the mixture NTCP model. The advantages and disadvantages of three main approaches (comparison of delineated volumes, overlap measures and Hausdorff distance) are discussed below.

Comparison of delineated volumes is the most common approach for assessing variability in organ delineation [143]. A limitation of this method is that it gives no information about overlap: two structures might have the same volume, but different locations. In this study, the variability of delineated $V_{thyroid}$, D_{mean} and estimated risk of RIHT were assessed using Bland-Altman plots. These plots allow visual assessment of measurement error (both systematic and random) of continuous variables [113,144].

Overlap measures take both volume and location into account. Two commonly used approaches are the Sørensen-Dice Similarity index (DSI) and the Jacard similarity index (JSI) (also called concordance index) [145]. The definitions of the two indexes are different but related; both have the value of 1 when there is perfect overlap between two structures, and a value of 0 when there is no overlap. However, the index for the same partial overlap is different, and the interpretation of values between 0 and 1 is not straightforward. DSI has been shown to be a special type of kappa statistics [146], and thus, a DSI >0.7 is commonly referred to as good agreement. We calculated DSI to facilitate comparison with other studies. DSI is a relative measure, however, and large structures are more likely to have higher DSI than small structures with the same variability [147,148]. Comparisons of DSI values should thus be limited to delineation studies of one specific organ. Another limitation of overlap measures is that they do not provide information on how contours may vary quantitatively in size, shape or location in absolute terms [143].

Hausdorff distance is often used to assess variation between delineations. For two surfaces, Hausdorff distance is defined as the maximum distance from any point on one surface to the closest point on the other surface [149]. It does not describe local variations, however, and there is no standard method for assessing local variability between delineations. Different 3D models of surface variation can thus be used. Brouwer et al. [150] assessed the local variation in five distances to the median (SD) of the delineations to obtain a model of global 3D SD. We developed a method for displaying both systematic deviations (mean values) and random variations (SD) in local variation, using the center of mass as origin and measuring the distance between the contours.

Strengths of the study were inclusion of a considerable sample size, judged to be representative for the population included in

Study I, and the use of different methods to study the variability in thyroid gland delineation.
A **limitation** was the low number of observers. This was addressed in Paper II by analyzing the impact of extreme systematic variations on the estimated NTCP. Despite this, we cannot rule out an impact of local delineation traditions or training as the reason for the small systematic variation between delineations.

Study III

The aim of this study was to validate the mixture NTCP model developed in Study I (cohort1, OUH) in an equivalent cohort treated at a different hospital (cohort2, AUH). Patients at both hospitals were treated according to DAHANCA guidelines and the same inclusion and exclusion criteria were applied in both studies.
It became evident during analysis that there were some differences between the two cohorts that had not been considered as a limitation during planning of the study. While the difference in number of TSH assessments for two cohorts was adjusted for by the sampling of cohort1S, latency could not be estimated due to very large uncertainties on the estimates. Bentzen et al. [112] have previously estimated latency in late skin reactions from a single follow-up, hence the assumption that this would be possible in cohort2. Both the sampling of a new cohort and the use of a fixed time factor from Study I might have been avoided by planning additional follow-ups and a study period to obtain more TSH assessments.
Baseline TSH was a significant risk factor in cohort2, but was not included in the model because we wanted to study an NTCP model equivalent to our model from cohort1, which only included $V_{thyroid}$ and D_{mean}.
A **strength** of the study was the inclusion of a similar size cohort as in Study I. As we used the same inclusion and exclusion criteria as in Study I, and included consecutive patients from the outpatient clinic, we believe that the cohort is representative for patients treated with definitive radiotherapy at AUH. The blood samples were analyzed in the same laboratory in Studies I and III.
Limitations of the study were the partly retrospective design and the limited number of follow-up TSH assessments (as discussed above). A bias in baseline TSH due to the timing of blood sampling is conceivable, since most of the patients had their baseline TSH during the first two weeks of radiotherapy. At this time the radiotherapy may theoretically have caused thyroiditis, which could have affected (typically lowering) the TSH levels.

CONCLUSION AND PERSPECTIVES

Radiation-induced hypothyroidism (RIHT) is a frequent late effect after definitive radiotherapy for head and neck squamous cell carcinoma. Dose constrains for the thyroid gland should thus be applied in radiation treatment planning to avoid RIHT, without compromising target coverage. D_{mean} and $V_{thyroid}$ were significant risk factors for RIHT in two independent patient cohorts. Both factors were included in the NTCP models developed in Study I and III, that provide suggested dose constraints in relation to specific thyroid volumes.
Validation of the input data for the NTCP1 model (Study II) showed it to be robust against observer variations in delineations of the thyroid gland at the population level. There may be pronounced differences in estimated individual risk of RIHT, however. Precise delineation is thus of utmost importance.
The NTCP1 model could not be fully validated and could not be directly compared to NTCP2. However, when the number of TSH assessments was taken into account, the estimated risks were very similar.
In light of the frequent occurrence of RIHT, and the potential repercussions of leaving this untreated, we recommend routine follow-up of thyroid function with regular TSH assessments after radiotherapy to the head and neck region. Since RIHT predominantly developed within 2-3 years after radiotherapy, routine TSH assessments in the first five years may suffice, and not lifelong as proposed by others [81,88,94].
The current studies have provided an NTCP model for RIHT and dose constraints for the thyroid gland that can be implemented in a clinical setting. Further validation of the model and dose constraints is desirable, however, in view of the uncertainties of NTCP models. Different strategies can be used for this. Despite the very similar risk estimates by NTCP1S and NTCP2, validation by obtaining more TSH assessments and a longer follow-up in cohort2 should be considered for validating the latency function and to form the basis for guidelines for follow-up assessment of thyroid function after radiotherapy. Pooling of data from more studies has been proposed for validation of NTCP models [151] as was done for establishing dose constraints for the parotid glands [119]. Pooling of the data from cohort1 and cohort2, and potentially data from other centers/countries with a longer follow-up, is an option that could be used to establish dose-response models on a very large cohort. To that end, uniform reporting of the data is needed. In support of this, we recommend biochemical hypothyroidism as an endpoint for modeling, with the possibility for further division of patients into subclinical and overt hypothyroidism.

The ultimate goal for further validation of the model is a prospective, longitudinal study with standardization of TSH assessments. Such a study could also include patient-reported outcomes, such as quality of life, in relation to RIHT. Furthermore, regular assessment of thyroid function might uncover the attributable factors and course of subclinical hypothyroidism after radiotherapy. Is it an inevitable road to overt hypothyroidism, or does remission occur?

ABBREVIATIONS

3D-CRT	Three dimensional conformal radiotherapy
AUH	Aarhus University Hospital
CRT	Conformal radiotherapy
CT	Computed tomography
DAHANCA	Danish Head and Neck Cancer Group
DICOM	Digital Imaging and Communications in
D_{mean}	Mean dose (Gy)
DSI	Sørensen-Dice similarity index
DVH	Dose volume histogram
HNSCC	Head and neck squamous cell carcinoma
HT	Hypothyroidism
IMRT	Intensity modulated radiotherapy
KBF	Department of Biochemistry and Pharmacology
NTCP	Normal tissue complication probability
OAR	Organs at risk
OUH	Odense University Hospital
RIHT	Radiation-induced hypothyroidism
RT	Radiotherapy
SD	Standard deviation
SHT	Subclinical hypothyroidism
TSH	Thyrotropin (thyroid stimulating hormone)
$V_{thyroid}$	Thyroid gland volume (cm^3)

SUMMARY

Head and neck cancer is the sixth most common cancer in the world. In Denmark the incidence was 1,287 in 2013 (thyroid cancer excluded). Head and neck cancer is predominantly a locoregional disease, and radiotherapy (RT) and surgery are major treatment options. Approximately 70% of Danish patients receive definitive RT, and the rest are treated with surgery with or without postoperative RT.

Radiation to the normal tissues is inevitable during RT. The current advanced treatment modalities allow precise calculation of radiation doses to normal tissues and there is a potential to distribute the dose to avoid adverse effects from radiation. However, this requires knowledge about the radiation tolerance levels of individual organs. Radiation-induced hypothyroidism (RIHT) is a well-known late effect of radiation to the thyroid gland, which can develop months to years after RT. The reported incidence of RIHT varies considerably, however, and the tolerance level of the thyroid gland is poorly defined. The aim of this PhD project was thus to elucidate these issues.

The first study included a cohort of 203 patients with head and neck cancer who were treated with definitive RT and no surgery at Odense University Hospital, Denmark, in 2002-2010. Analysis of patient characteristics, precise radiation doses to the thyroid gland and follow-up assessments of thyroid function after radiotherapy gave an estimated 5-year incidence of RIHT of 26%. Significant risk factors for the development of RIHT were a small thyroid volume and a high mean radiation dose to the gland. Two models were developed for predicting the risk of RIHT, taking both the thyroid volume and the mean radiation dose into account. One model was based on the collected data alone, while the other model included an estimated time distribution for the development of RIHT (i.e. latency). Using the model that considered latency, dose constraints for the thyroid gland were proposed for keeping the risk of RIHT after radiotherapy below 25%. This model emerged as the preferred model, and its robustness was tested in the following two studies.

The second study aimed at assessing the impact of variation of estimated risk of RIHT due to intra- and inter-observer variability in delineated thyroid gland volume. Fifty treatment plans were randomly chosen for either repeat delineation by the author (intra-observer) or delineation by another radiation oncologist (inter-observer). The variations in delineated thyroid gland volume and mean radiation dose, and the subsequent variation in estimated risk of RIHT were assessed. For the entire study population, the variation in predicted risk of RIHT was small and the model was robust towards observer variations in delineation of the thyroid gland. However, for some patients there were pronounced differences in the estimated risk due to variation in organ delineation, and thus precise delineation is of utmost importance for the individual patient.

The third study aimed at studying the robustness of the model in a cohort of head and neck cancer patients treated with definitive RT at the Department of Oncology, Aarhus University Hospital. 198 patients were included after one assessment of thyroid function (blood sample of thyrotropin, TSH) during routine follow-up after radiotherapy. A small thyroid volume and a high mean radiation dose were also found to be significant risk factors for radiation-induced hypothyroidism in this cohort. These factors were included in a new risk model. Direct comparison of the two models was not possible, however, due to a different number of TSH assessments in the two cohorts. It was also not possible to estimate latency on the basis of a single measurement of thyroid function. Nevertheless, when the number of TSH assessments was taken into account, the estimated risk of RIHT was very similar in the two cohorts.

Conclusion

Hypothyroidism is a frequent late effect after definitive radiotherapy. As the condition has been linked to increased risk of cardiac disease and mortality, and decreased quality of life, it is important to consider the risk of radiation-induced hypothyroidism (RIHT) when planning radiation treatment. The size of the thyroid gland and the radiation dose to the gland are key factors in the development of RIHT, and both these factors should be considered when determining dose constraints for the thyroid gland. A risk of RIHT below 25% is recommended. Furthermore, routine assessment of thyroid gland function should be offered after radiotherapy in the neck area. Finally, the robustness of the risk model should be studied in a longitudinal follow-up of a cohort of patients with repeated thyroid function assessment.

REFERENCES

1. Boyle P, Levine B, editors. World Cancer Report 2008.
2. DAHANCA. Årsrapport 2013 for den kliniske kvalitetsdatabase DAHANCA [Annual report for the DAHANCA database]2014.
3. Statens Serum Institut. Cancerregisteret. Tal og analyse. [The Danish Cancer Register. Numbers and analysis]. www.ssi.dk2014.
4. Statens Serum Institut. Kræftoverlevelse i Danmark 1998-2012.[Cancer survival in Denmark 1998-2012]. www.ssi.dk2014.
5. Lassen P. The role of Human papillomavirus in head and neck cancer and the impact on radiotherapy outcome. Radiother Oncol 2010;95:371-80.
6. Garnaes E, Kiss K, Andersen L, et al. A high and increasing HPV prevalence in tonsillar cancers in Eastern Denmark, 2000-2010: The largest registry-based study to date. Int J Cancer 2014.
7. Lassen P, Eriksen JG, Krogdahl A, et al. The influence of HPV-associated p16-expression on accelerated fractionated radiotherapy in head and neck cancer: evaluation of the randomised DAHANCA 6&7 trial. Radiother Oncol 2011;100:49-55.
8. Bol V, Gregoire V. Biological basis for increased sensitivity to radiation therapy in HPV-positive head and neck cancers. Biomed Res Int 2014;2014:696028.
9. Overgaard J, Hansen HS, Jorgensen K, Hjelm Hansen M. Primary radiotherapy of larynx and pharynx carcinoma--an analysis of some factors influencing local control and survival. Int J Radiat Oncol Biol Phys 1986;12:515-21.
10. Blanchard P, Baujat B, Holostenco V, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. Radiother Oncol 2011;100:33-40.
11. Overgaard J, Hansen HS, Overgaard M, et al. A randomized double-blind phase III study of nimorazole as a hypoxic radiosensitizer of primary radiotherapy in supraglottic larynx and pharynx carcinoma, results of the Danish Head and Neck Cancer Study (DAHANCA) protocol 5-85. Radiother Oncol 1998;46:135-46.
12. Overgaard J, Hansen HS, Specht L, et al. Five compared with six fractions per week of conventional radiotherapy of squa-

mous-cell carcinoma of head and neck: DAHANCA 6 and 7 randomised controlled trial. Lancet 2003;362:933-40.
13. Bourhis J, Overgaard J, Audry H, et al. Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. Lancet 2006;368:843-54.
14. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. N Engl J Med 2006;354:567-78.
15. Bonner JA, Harari PM, Giralt J, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. Lancet Oncol 2010;11:21-8.
16. Ang KK, Zhang Q, Rosenthal DI, et al. Randomized phase III trial of concurrent accelerated radiation plus cisplatin with or without cetuximab for stage III to IV head and neck carcinoma: RTOG 0522. J Clin Oncol 2014;32:2940-50.
17. Eriksen JG, Maare C, Johansen J, et al. Evaluation of the EGFR-Inhibitor zalutumumab given with primary curative (CHEMO) radiation therapy to patients with squamous cell carcinoma of the head and neck: Results of the DAHANCA 19 randomized phase 3 trial. Int J Radiat Oncol Biol Phys 2014;88 (2):465.
18. Marta GN, Silva V, de Andrade Carvalho H, et al. Intensity-modulated radiation therapy for head and neck cancer: systematic review and meta-analysis. Radiother Oncol 2014;110:9-15.
19. Nutting CM, Morden JP, Harrington KJ, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol 2011;12:127-36.
20. Bertelsen A, Hansen CR, Johansen J, Brink C. Single Arc Volumetric Modulated Arc Therapy of head and neck cancer. Radiother Oncol 2010;95:142-8.
21. Danish Head and Neck Cancer Group (DAHANCA) guidelines. 2013. (Accessed cited 2014.10.28, at www.dahanca.dk/guidelines.)
22. Overgaard J, Hansen HS, Andersen AP, et al. Misonidazole combined with split-course radiotherapy in the treatment of invasive carcinoma of larynx and pharynx: report from the DAHANCA 2 study. Int J Radiat Oncol Biol Phys 1989;16:1065-8.
23. Boje CR, Dalton SO, Gronborg TK, et al. The impact of comorbidity on outcome in 12 623 Danish head and neck cancer patients: a population based study from the DAHANCA database. Acta Oncol 2013;52:285-93.
24. Hansen C, Johansen J, Kristensen CA, et al. National audit of DAHANCA radiation treatment protocols. Radiotherapy and Oncology 2010;96:S323-S4.
25. Boje CR, Dalton SO, Primdahl H, et al. Evaluation of comorbidity in 9388 head and neck cancer patients: a national cohort study from the DAHANCA database. Radiother Oncol 2014;110:91-7.
26. Pagh A, Vedtofte T, Lynggaard CD, et al. The value of routine follow-up after treatment for head and neck cancer. A national survey from DAHANCA. Acta Oncol 2013;52:277-84.
27. Toustrup K, Sorensen BS, Lassen P, Wiuf C, Alsner J, Overgaard J. Gene expression classifier predicts for hypoxic modification of radiotherapy with nimorazole in squamous cell carcinomas of the head and neck. Radiother Oncol 2012;102:122-9.
28. Mortensen HR, Jensen K, Aksglaede K, Behrens M, Grau C. Late dysphagia after IMRT for head and neck cancer and correlation with dose-volume parameters. Radiother Oncol 2013;107:288-94.
29. Lassen P, Primdahl H, Johansen J, et al. Impact of HPV-associated p16-expression on radiotherapy outcome in advanced oropharynx and non-oropharynx cancer. Radiother Oncol 2014;113:310-6.
30. Internaltional Commission on Radiation Units & Measurements. ICRU Report 83. Chapter 5. Planning Aims, Prescriptions, and Technical Data. Journal of the ICRU2010.
31. Trotti A, Colevas AD, Setser A, et al. CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. Semin Radiat Oncol 2003;13:176-81.
32. Mortensen HR, Overgaard J, Specht L, et al. Prevalence and peak incidence of acute and late normal tissue morbidity in the DAHANCA 6&7 randomised trial with accelerated radiotherapy for head and neck cancer. Radiother Oncol 2012;103:69-75.
33. Follow-up. 2000. at http://www.dahanca.dk/get_media_file.php?mediaid=100.)
34. Pavy JJ, Denekamp J, Letschert J, et al. EORTC Late Effects Working Group. Late effects toxicity scoring: the SOMA scale. Radiother Oncol 1995;35:11-5.
35. Rubin P, Constine LS, Fajardo LF, Phillips TL, Wasserman TH. RTOG Late Effects Working Group. Overview. Late Effects of Normal Tissues (LENT) scoring system. Int J Radiat Oncol Biol Phys 1995;31:1041-2.
36. Bentzen SM, Dorr W, Anscher MS, et al. Normal tissue effects: reporting and analysis. Semin Radiat Oncol 2003;13:189-202.
37. Bentzen SM. Basic Clinical Radiobiology. Chapter 5. Dose-response relationships in radiotherapy. Fourth edition ed. London: Hodder Arnold; 2009.
38. Bentzen SM, Tucker SL. Quantifying the position and steepness of radiation dose-response curves. Int J Radiat Biol 1997;71:531-42.
39. Bentzen SM, Thames HD, Travis EL, et al. Direct estimation of latent time for radiation injury in late-responding normal tissues: gut, lung, and spinal cord. Int J Radiat Biol 1989;55:27-43.
40. Emami B, Lyman J, Brown A, et al. Tolerance of normal tissue to therapeutic irradiation. Int J Radiat Oncol Biol Phys 1991;21:109-22.
41. Bentzen SM, Constine LS, Deasy JO, et al. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. Int J Radiat Oncol Biol Phys 2010;76:S3-9.
42. Hegedüs L, Perrild H, Poulsen LR, et al. The determination of thyroid volume by ultrasound and its relationship to body weight, age, and sex in normal subjects. J Clin Endocrinol Metab 1983;56:260-3.
43. Krejbjerg A, Bjergved L, Pedersen IB, et al. Iodine fortification may influence the age-related change in thyroid volume: a longitudinal population-based study (DanThyr). Eur J Endocrinol 2014;170:507-17.
44. Hansen PS, Brix TH, Bennedbaek FN, Bonnema SJ, Kyvik KO, Hegedüs L. Genetic and environmental causes of individual differences in thyroid size: a study of healthy Danish twins. J Clin Endocrinol Metab 2004;89:2071-7.
45. Weeke J, Gundersen HJ. Circadian and 30 minutes variations in serum TSH and thyroid hormones in normal subjects. Acta Endocrinol 1978;89:659-72.

46. Brabant G, Prank K, Ranft U, et al. Physiological regulation of circadian and pulsatile thyrotropin secretion in normal man and woman. J Clin Endocrinol Metab 1990;70:403-9.
47. Hancock SL, McDougall IR, Constine LS. Thyroid abnormalities after therapeutic external radiation. Int J Radiat Oncol Biol Phys 1995;31:1165-70.
48. Jereczek-Fossa BA, Alterio D, Jassem J, Gibelli B, Tradati N, Orecchia R. Radiotherapy-induced thyroid disorders. Cancer Treat Rev 2004;30:369-84.
49. Bakhshandeh M, Hashemi B, Mahdavi SR, Nikoofar A, Edraki HR, Kazemnejad A. Evaluation of thyroid disorders during head-and-neck radiotherapy by using functional analysis and ultrasonography. Int J Radiat Oncol Biol Phys 2012;83:198-203.
50. Lo Galbo AM, Kuik DJ, Lips P, et al. A prospective longitudinal study on endocrine dysfunction following treatment of laryngeal or hypopharyngeal carcinoma. Oral Oncol 2013;49:950-5.
51. Carr RF, LiVolsi VA. Morphologic changes in the thyroid after irradiation for Hodgkin's and non-Hodgkin's lymphoma. Cancer 1989;64:825-9.
52. Jensen E, Hyltoft Petersen P, Blaabjerg O, et al. Establishment of a serum thyroid stimulating hormone (TSH) reference interval in healthy adults. The importance of environmental factors, including thyroid antibodies. Clin Chem Lab Med 2004;42:824-32.
53. Duntas LH, Brenta G. The effect of thyroid disorders on lipid levels and metabolism. Med Clin North Am 2012;96:269-81.
54. Erem C. Thyroid disorders and hypercoagulability. Semin Thromb Hemost 2011;37:17-26
55. Lekakis J, Papamichael C, Alevizaki M, et al. Flow-mediated, endothelium-dependent vasodilation is impaired in subjects with hypothyroidism, borderline hypothyroidism, and high-normal serum thyrotropin (TSH) values. Thyroid 1997;7:411-4.
56. Nyirenda MJ, Clark DN, Finlayson AR, et al. Thyroid disease and increased cardiovascular risk. Thyroid 2005;15:718-24.
57. Biondi B, Cooper DS. The clinical significance of subclinical thyroid dysfunction. Endocr Rev 2008;29:76-131.
58. Thvilum M, Brandt F, Almind D, Christensen K, Brix TH, Hegedüs L. Type and extent of somatic morbidity before and after the diagnosis of hypothyroidism. a nationwide register study. PLoS One 2013;8:e75789.
59. Thvilum M, Brandt F, Almind D, Christensen K, Brix TH, Hegedüs L. Increased psychiatric morbidity before and after the diagnosis of hypothyroidism: a nationwide register study. Thyroid 2014;24:802-8.
60. Watt T, Groenvold M, Rasmussen AK, et al. Quality of life in patients with benign thyroid disorders. A review. Eur J Endocrinol 2006;154:501-10.
61. Watt T, Cramon P, Hegedüs L, et al. The thyroid-related quality of life measure ThyPRO has good responsiveness and ability to detect relevant treatment effects. J Clin Endocrinol Metab 2014;99:3708-17.
62. Thvilum M, Brandt F, Brix TH, Hegedüs L. A review of the evidence for and against increased mortality in hypothyroidism. Nat Rev Endocrinol 2012;8:417-24.
63. Thvilum M, Brandt F, Almind D, Christensen K, Hegedüs L, Brix TH. Excess mortality in patients diagnosed with hypothyroidism: a nationwide cohort study of singletons and twins. J Clin Endocrinol Metab 2013;98:1069-75.
64. Laulund AS, Nybo M, Brix TH, Abrahamsen B, Jorgensen HL, Hegedüs L. Duration of Thyroid Dysfunction Correlates with All-Cause Mortality. The OPENTHYRO Register Cohort. PLoS One 2014;9:e110437.
65. Rodondi N, Aujesky D, Vittinghoff E, Cornuz J, Bauer DC. Subclinical hypothyroidism and the risk of coronary heart disease: a meta-analysis. Am J Med 2006;119:541-51.
66. Selmer C, Olesen JB, Hansen ML, et al. Subclinical and overt thyroid dysfunction and risk of all-cause mortality and cardiovascular events: a large population study. J Clin Endocrinol Metab 2014;99:2372-82.
67. Singh S, Duggal J, Molnar J, Maldonado F, Barsano CP, Arora R. Impact of subclinical thyroid disorders on coronary heart disease, cardiovascular and all-cause mortality: a meta-analysis. Int J Cardiol 2008;125:41-8.
68. Asvold BO, Bjoro T, Platou C, Vatten LJ. Thyroid function and the risk of coronary heart disease: 12-year follow-up of the HUNT study in Norway. Clin Endocrinol 2012;77:911-7.
69. Rodondi N, den Elzen WP, Bauer DC, et al. Subclinical hypothyroidism and the risk of coronary heart disease and mortality. JAMA 2010;304:1365-74.
70. Collet TH, Bauer DC, Cappola AR, et al. Thyroid antibody status, subclinical hypothyroidism, and the risk of coronary heart disease: an individual participant data analysis. J Clin Endocrinol Metab 2014;99:3353-62.
71. McQuade C, Skugor M, Brennan DM, Hoar B, Stevenson C, Hoogwerf BJ. Hypothyroidism and moderate subclinical hypothyroidism are associated with increased all-cause mortality independent of coronary heart disease risk factors: a PreCIS database study. Thyroid 2011;21:837-43.
72. Vanderpump MP, Tunbridge WM. Epidemiology and prevention of clinical and subclinical hypothyroidism. Thyroid 2002;12:839-47.
73. Carle A, Laurberg P, Pedersen IB, et al. Epidemiology of subtypes of hypothyroidism in Denmark. Eur J Endocrinol 2006;154:21-8.
74. Knudsen N, Bulow I, Jorgensen T, Laurberg P, Ovesen L, Perrild H. Comparative study of thyroid function and types of thyroid dysfunction in two areas in Denmark with slightly different iodine status. Eur J Endocrinol 2000;143:485-91.
75. Pearce SH, Brabant G, Duntas LH, et al. 2013 ETA Guideline: Management of Subclinical Hypothyroidism. Eur Thyroid J 2013;2:215-28.
76. Brix TH, Kyvik KO, Hegedüs L. A population-based study of chronic autoimmune hypothyroidism in Danish twins. J Clin Endocrinol Metab 2000;85:536-9.
77. Brix TH, Hansen PS, Kyvik KO, Hegedüs L. Cigarette smoking and risk of clinically overt thyroid disease: a population-based twin case-control study. Arch Intern Med 2000;160:661-6.
78. Laurberg P, Cerqueira C, Ovesen L, et al. Iodine intake as a determinant of thyroid disorders in populations. Best Pract Res Clin Endocrinol Metab 2010;24:13-27.
79. Brix TH, Hegedüs L. Twin studies as a model for exploring the aetiology of autoimmune thyroid disease. Clin Endocrinol 2012;76:457-64.
80. Tell R, Sjodin H, Lundell G, Lewin F, Lewensohn R. Hypothyroidism after external radiotherapy for head and neck cancer. Int J Radiat Oncol Biol Phys 1997;39:303-8.
81. Tell R, Lundell G, Nilsson B, Sjodin H, Lewin F, Lewensohn R. Long-term incidence of hypothyroidism after radiotherapy in patients with head-and-neck cancer. Int J Radiat Oncol Biol Phys 2004;60:395-400.

82. Kumpulainen EJ, Hirvikoski PP, Virtaniemi JA, et al. Hypothyroidism after radiotherapy for laryngeal cancer. Radiother Oncol 2000;57:97-101.
83. Sinard RJ, Tobin EJ, Mazzaferri EL, et al. Hypothyroidism after treatment for nonthyroid head and neck cancer. Arch Otolaryngol Head Neck Surg 2000;126:652-7.
84. Hancock SL, Cox RS, McDougall IR. Thyroid diseases after treatment of Hodgkin's disease. N Engl J Med 1991;325:599-605.
85. Bhandare N, Kennedy L, Malyapa RS, Morris CG, Mendenhall WM. Primary and central hypothyroidism after radiotherapy for head-and-neck tumors. Int J Radiat Oncol Biol Phys 2007;68:1131-9.
86. Norris AA, Amdur RJ, Morris CG, Mendenhall WM. Hypothyroidism when the thyroid is included only in the low neck field during head and neck radiotherapy. Am J Clin Oncol 2006;29:442-5.
87. Turner SL, Tiver KW, Boyages SC. Thyroid dysfunction following radiotherapy for head and neck cancer. Int J Radiat Oncol Biol Phys 1995;31:279-83.
88. Mercado G, Adelstein DJ, Saxton JP, Secic M, Larto MA, Lavertu P. Hypothyroidism: a frequent event after radiotherapy and after radiotherapy with chemotherapy for patients with head and neck carcinoma. Cancer 2001;92:2892-7.
89. Lo Galbo AM, de Bree R, Kuik DJ, et al. The prevalence of hypothyroidism after treatment for laryngeal and hypopharyngeal carcinomas: are autoantibodies of influence? Acta Otolaryngol 2007;127:312-7.
90. August M, Wang J, Plante D, Wang CC. Complications associated with therapeutic neck radiation. J Oral Maxillofac Surg 1996;54:1409-15; discussion 15-6.
91. Diaz R, Jaboin JJ, Morales-Paliza M, et al. Hypothyroidism as a consequence of intensity-modulated radiotherapy with concurrent taxane-based chemotherapy for locally advanced head-and-neck cancer. Int J Radiat Oncol Biol Phys 2010;77:468-76.
92. Colevas AD, Read R, Thornhill J, et al. Hypothyroidism incidence after multimodality treatment for stage III and IV squamous cell carcinomas of the head and neck. Int J Radiat Oncol Biol Phys 2001;51:599-604.
93. Kuten A, Lubochitski R, Fishman G, Dale J, Stein ME. Postradiotherapy hypothyroidism: Radiation dose response and chemotherapeutic radiosensitization at less than 40 Gy. Journal of Surgical Oncology 1996;61:281-3.
94. Garcia-Serra A, Amdur RJ, Morris CG, Mazzaferri E, Mendenhall WM. Thyroid function should be monitored following radiotherapy to the low neck. Am J Clin Oncol 2005;28:255-8.
95. Grande C. Hypothyroidism following radiotherapy for head and neck cancer: multivariate analysis of risk factors. Radiother Oncol 1992;25:31-6.
96. Alterio D, Jereczek-Fossa BA, Franchi B, et al. Thyroid disorders in patients treated with radiotherapy for head-and-neck cancer: a retrospective analysis of seventy-three patients. Int J Radiat Oncol Biol Phys 2007;67:144-50.
97. Wu YH, Wang HM, Chen HH, et al. Hypothyroidism after radiotherapy for nasopharyngeal cancer patients. Int J Radiat Oncol Biol Phys 2010;76:1133-9.
98. Lin Z, Wu VW, Lin J, Feng H, Chen L. A longitudinal study on the radiation-induced thyroid gland changes after external beam radiotherapy of nasopharyngeal carcinoma. Thyroid 2011;21:19-23.
99. Cella L, Conson M, Caterino M, et al. Thyroid V30 predicts radiation-induced hypothyroidism in patients treated with sequential chemo-radiotherapy for Hodgkin's lymphoma. Int J Radiat Oncol Biol Phys 2012;82:1802-8.
100. Cella L, Liuzzi R, Conson M, D'Avino V, Salvatore M, Pacelli R. Development of multivariate NTCP models for radiation-induced hypothyroidism: a comparative analysis. Radiat Oncol 2012;7:224.
101. Boomsma MJ, Bijl HP, Christianen ME, et al. A prospective cohort study on radiation-induced hypothyroidism: development of an NTCP model. Int J Radiat Oncol Biol Phys 2012;84:e351-6.
102. Bakhshandeh M, Hashemi B, Mahdavi SR, Nikoofar A, Vasheghani M, Kazemnejad A. Normal tissue complication probability modeling of radiation-induced hypothyroidism after head-and-neck radiation therapy. Int J Radiat Oncol Biol Phys 2013;85:514-21.
103. Rønjom MF, Brink C, Bentzen SM, Hegedüs L, Overgaard J, Johansen J. Hypothyroidism after primary radiotherapy for head and neck squamous cell carcinoma: normal tissue complication probability modeling with latent time correction. Radiother Oncol 2013;109:317-22.
104. Lin Z, Wang X, Xie W, Yang Z, Che K, Wu VW. Evaluation of clinical hypothyroidism risk due to irradiation of thyroid and pituitary glands in radiotherapy of nasopharyngeal cancer patients. J Med Imaging Radiat Oncol 2013;57:713-8.
105. Lin Z, Chen L, Fang Y, Cai A, Zhang T, Wu VW. Longitudinal study on the correlations of thyroid antibody and thyroid hormone levels after radiotherapy in patients with nasopharyngeal carcinoma with radiation-induced hypothyroidism. Head Neck 2014;36:171-5.
106. Murthy V, Narang K, Ghosh-Laskar S, Gupta T, Budrukkar A, Agrawal JP. Hypothyroidism after 3-dimensional conformal radiotherapy and intensity-modulated radiotherapy for head and neck cancers: Prospective data from 2 randomized controlled trials. Head Neck 2014;36:1573-80.
107. Akgun Z, Atasoy BM, Ozen Z, et al. V30 as a predictor for radiation-induced hypothyroidism: a dosimetric analysis in patients who received radiotherapy to the neck. Radiat Oncol 2014;9:104.
108. Kim MY, Yu T, Wu HG. Dose-volumetric Parameters for Predicting Hypothyroidism after Radiotherapy for Head and Neck Cancer. Jpn J Clin Oncol 2014.
109. Chyan A, Chen J, Shugard E, Lambert L, Quivey JM, Yom SS. Dosimetric predictors of hypothyroidism in oropharyngeal cancer patients treated with intensity-modulated radiation therapy. Radiat Oncol 2014;9:269.
110. Sachdev S, Refaat T, Bacchus ID, Sathiaseelan V, Mittal BB. Thyroid V50 Highly Predictive of Hypothyroidism in Head-and-Neck Cancer Patients Treated With Intensity-modulated Radiotherapy (IMRT). Am J Clin Oncol 2014.
111. Protocol DAHANCA 10 study. 2003. (Accessed 2014.11.28, at http://www.dahanca.dk/get_media_file.php?mediaid=43.)
112. Bentzen SM, Thames HD, Overgaard M. Latent-time estimation for late cutaneous and subcutaneous radiation reactions in a single-follow-up clinical study. Radiother Oncol 1989;15:267-74.
113. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
114. Westberg J, Krogh S, Brink C, Vogelius IR, Iop. A DICOM based radiotherapy plan database for research collaboration and reporting. Xvii International Conference on the Use of Computers in Radiation Therapy (Iccr 2013) 2014;489:5.

115. Vogelius IR, Bentzen SM, Maraldo MV, Petersen PM, Specht L. Risk factors for radiation-induced hypothyroidism: a literature-based meta-analysis. Cancer 2011;117:5250-60.
116. Johansen J, Overgaard J. Årsberetning 2013. Den danske Hoved-Halscancer Gruppe (DAHANCA), Dansk Multidicipli-nære Cancer Grupper (DMCG.dk)2014.
117. Lawrence YR, Li XA, el Naqa I, et al. Radiation dose-volume effects in the brain. Int J Radiat Oncol Biol Phys 2010;76:S20-7.
118. Darzy KH, Shalet SM. Hypopituitarism following radiotherapy. Pituitary 2009;12:40-50.
119. Dijkema T, Raaijmakers CP, Ten Haken RK, et al. Parotid gland function after radiotherapy: the combined michigan and utrecht experience. Int J Radiat Oncol Biol Phys 2010;78:449-53.
120. Deasy JO, Moiseenko V, Marks L, Chao KS, Nam J, Eisbruch A. Radiotherapy dose-volume effects on salivary gland function. Int J Radiat Oncol Biol Phys 2010;76:S58-63.
121. Christianen ME, Schilstra C, Beetz I, et al. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study. Radiother Oncol 2012;105:107-14.
122. Rancati T, Schwarz M, Allen AM, et al. Radiation dose-volume effects in the larynx and pharynx. Int J Radiat Oncol Biol Phys 2010;76:S64-9.
123. Policeni BA, Smoker WR, Reede DL. Anatomy and embryology of the thyroid and parathyroid glands. Semin Ultrasound CT MR 2012;33:104-14.
124. Bashist B, Ellis K, Gold RP. Computed tomography of intrathoracic goiters. AJR Am J Roentgenol 1983;140:455-60.
125. Kaneko T, Matsumoto M, Fukui K, Hori T, Katayama K. Clinical evaluation of thyroid CT values in various thyroid conditions. J Comput Tomogr 1979;3:1-4.
126. Marks LB, Yorke ED, Jackson A, et al. Use of normal tissue complication probability models in the clinic. Int J Radiat Oncol Biol Phys 2010;76:S10-9.
127. U. S. Department of Health and Human Services. Common Terminology Criteria for Adverse Events (CTCAE). v4.03.2009 June 14, 2010.
128. LENT SOMA tables. Radiother Oncol 1995;35:17-60.
129. The Biologic Variation database. Database update. 2014. at https://www.westgard.com/biodatabase1.htm.)
130. Ricos C, Alvarez V, Cava F, et al. Current databases on biological variation: pros, cons and progress. Scand J Clin Lab Invest 1999;59:491-500.
131. Perich C, Minchinela J, Ricos C, et al. Biological variation database: structure and criteria used for generation and update. Clin Chem Lab Med 2014.
132. Andersen S, Bruun NH, Pedersen KM, Laurberg P. Biologic variation is important for interpretation of thyroid function tests. Thyroid 2003;13:1069-78.
133. Jensen E, Blaabjerg O, Petersen PH, Hegedüs L. Sampling time is important but may be overlooked in establishment and use of thyroid-stimulating hormone reference intervals. Clin Chem 2007;53:355-6.
134. Laurberg P, Andersen S, Carle A, Karmisholt J, Knudsen N, Pedersen IB. The TSH upper reference limit: where are we at? Nat Rev Endocrinol 2011;7:232-9.
135. Barbesino G. Drugs affecting thyroid function. Thyroid 2010;20:763-70.
136. Feldt-Rasmussen U, Hyltoft Petersen P, Blaabjerg O, Horder M. Long-term variability in serum thyroglobulin and thyroid related hormones in healthy subjects. Acta Endocrinol (Copenh) 1980;95:328-34.
137. Andersen S, Pedersen KM, Bruun NH, Laurberg P. Narrow individual variations in serum T(4) and T(3) in normal subjects: a clue to the understanding of subclinical thyroid disease. J Clin Endocrinol Metab 2002;87:1068-72.
138. Withers HR, Taylor JM, Maciejewski B. Treatment volume and tissue tolerance. Int J Radiat Oncol Biol Phys 1988;14:751-9.
139. Seppenwoolde Y, Lebesque JV, de Jaeger K, et al. Comparing different NTCP models that predict the incidence of radiation pneumonitis. Int J Radiat Oncol Biol Phys 2003;55:724-35.
140. Pan CC, Kavanagh BD, Dawson LA, et al. Radiation-associated liver injury. Int J Radiat Oncol Biol Phys 2010;76:S94-100.
141. Kirkwood BR, Sterne AC. Essential Medical Statistics. Chapter 30. Second edition ed. Oxford: Blackwell Publishing; 2003.
142. Petruccelli JD, Nandram B, Chen M. Applied Statistics for Engineers and Scientists: Prentice Hall; 1999.
143. Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. J Med Imaging Radiat Oncol 2010;54:401-10.
144. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. A Practical Guide: Cambridge; 2011.
145. Lorenzen EL, Taylor CW, Maraldo M, et al. Inter-observer variation in delineation of the heart and left anterior descending coronary artery in radiotherapy for breast cancer: a multi-centre study from Denmark and the UK. Radiother Oncol 2013;108:254-8.
146. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Med Imaging 1994;13:716-24.
147. Zou KH, Warfield SK, Bharatha A, et al. Statistical validation of image segmentation quality based on a spatial overlap index. Acad Radiol 2004;11:178-89.
148. Rohlfing T, Brandt R, Menzel R, Russakoff DB, Calvin R. Maurer J. Quo Vadis, Atlas-Based Segmentation? In: Suri J, Wilson DL, Laxminarayan S, eds. The Handbook of Medical Image Analysis: Registration Models: Kluwer Academic/Plenum Publishers. New York, NY; 2005:435-86.
149. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing Images Using the Hausdorff Distance. IEEE Trans Med Imaging 1993;15:850-63.
150. Brouwer CL, Steenbakkers RJ, van den Heuvel E, et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol 2012;7:32.
151. Deasy JO, Bentzen SM, Jackson A, et al. Improving normal tissue complication probability models: the need to adopt a "data-pooling" culture. Int J Radiat Oncol Biol Phys 2010;76:S151-4.