University of Oslo

Management of Mild Primary Hyperparathyroidism

Dissertation for the Degree of PhD

Mikkel Pretorius – 2023

Section of Specialized Endocrinology

Division of Medicine, Oslo University Hospital and

Faculty of Medicine, University of Oslo

Oslo, Norway

2023

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-348-0226-3

Cover: UiO.
Print production: Graphic Center, University of Oslo.

Table of Content

Acknowledgements

First, I would like to thank the 191 patients included in the SIPH study. They agreed to be randomly selected for neck surgery and obeying by this result for a decade of investigations without obvious advantages for themselves. Their willingness and patience have made these publications and this thesis possible.

Secondly, I would like to express my gratitude and admiration for my supervisors. My principal supervisor Ansgar Heck has been an essential part of this journey. His helpfulness and availability in medical, scientific and practical issues have been vital in finishing this project. Thank you for facilitating the technical challenges in such a long-standing study as well as all our responsibilities towards the faculty. Your meticulous nature and dedication to science are something to strive towards.

Jens Bollerslev, co-supervisor and "founding father" of the SIHP study. His dedication to this project, as well as to his PhD-students, has been truly inspiring. Your warmth toward your colleagues and your coolness in times of turmoil is admirable. I feel truly blessed to have worked with one of the giants in the field of metabolic bone diseases.

Mikael Hellström, as co-supervisor, has been inspirational in both form and figure. His keen eye for the English language and scientific arguments have been decisive on more than one occasion. Your level of fitness, both physically and mentally, is impressive.

Karolina Lundstam, my fellow PhD-student that I owe so much. Your eye for detail and knowledge of the SIPH-study is impressive. Without your hard work and dedication to this study, this thesis would not have been possible. The scientific discussions and the talks about life outside of SIPH have been greatly appreciated.

Kristin Godang has been an administrative leader of SIPH and its data throughout my time in the study. Your countless hours organizing and validating patient data over decades have given us important insights into this common disease. Your knowledge of DXA and laboratory medicine is admirable and has been vital in all the publications.

Morten Wang Fagerland has been our statistical guide and mentor in all our publications. Your expertise and kind nature have been crucial for us as a research team and for me personally as a PhD-student.

The SIPH study crosses both decades and borders. I want to thank the entire SIPH-study group of Charlotte Mollerup, Stine L. Fougner, Ylva Pernow, Turid Aas, Ola Hessman, Thord Rosén, Jörgen Nordenström, and Svante Jansson, for your work with the patients and insightful comments on all the manuscripts.

Doing a research project in such a stimulating clinical and scientific environment has been enlightening. I thank all my colleagues at the Department of Endocrinology at Oslo University Hospital. To Anders Jørgensen, for your insights into the clinical field of endocrinology and continuous work to promote endocrinology in Norway. To Johan Arild Evang for all the laughs and for educating me in the important parts of life. To Kristin Astrid Øystese for sharing and caring. To Elisabeth Qvigstad for leading the research department with a steady hand. To Trine Finnes for making complicated things easy. To Kiarash Tazmini, for your positivity and enthusiasm. To Therese Weider for showing us there is more to life than endocrinology. To Kari Anne Sveen, for your empathy. To Tore Julsrud Berg for sound scientific advice along the way. To Sara Hammerstad for clinical and scientific expertise across the hall. I would especially like to thank my leader at Oslo University Hospital, Lene K. Seland, for making it possible to combine science with a clinical position at Oslo University Hospital and for your general support throughout all my years at OUS.

Thank you to South-Eastern Norway Regional Health Authority (Helse Sør-Øst RHF) for providing a research grant, making this PhD-project possible. Likewise, my gratitude includes the University of Oslo for facilitating mine and countless other PhD-projects.

Finally, I want to thank my family and friends. To my caring mother, Kine, who has always supported my projects and me. To my two brothers, Adrian and Jarl, for not focusing overly much on parathyroid hormone or science in general. My friends who have offered diversion and recreation during the tedious work of a PhD.
My deepest gratitude goes to my wife, Phuong Phuong. Your understanding, patience, and caring for our children and me have made this PhD-thesis possible and enjoyable. Your enthusiasm toward your own research project is contagious, and I am glad to share this journey of science and life with you. To my daughter Nora, your empathy and maturity is beyond years. You are wise, and I cherish the long talks and sound advice you give. To my son, Jonas, your enthusiasm and energy are abundant and infectious. Thank you for all the laughs and for reminding me to live in the moment.

Abbreviations

BMD	Bone mineral density
BMSi	Bone Material Strength index
CASR	Calcium-Sensing Receptor
CNS	Central Nervous System
CPRS	Comprehensive Psychopathological Rating Scale
DXA	Dual-energy X-ray absorptiometry
GDPR	General Data Protection Regulation
HR	Hazard Ratio
ITT	Intention to treat
IQOLA	International Quality of Life Assessment (IQOLA)
MCS	Mental Component Summary
NIH	National Institute for Health
PCS	Physical Component Summary
PHPT	Primary Hyperparathyroidism
PTH	Parathyroid Hormone
PROMs	Patient-Reported Outcome Measures
QoL	Quality of Life
RANKL	Receptor Activator Nuclear factor Kappa B Ligand
RCT	Randomized Controlled Trial
REC	Regional Ethical Committee
SD	Standard Deviation
SF-36	Short Form-36
SIPH	Scandinavian Study of Primary Hyperparathyroidism
TB	Total Body

Articles in the thesis:

Paper I: Effects of Parathyroidectomy on Quality of Life: 10 Years of Data From a Prospective Randomized Controlled Trial on Primary Hyperparathyroidism (the SIPH-Study)

Pretorius, M, Lundstam, K, Hellström, M, Fagerland, M.W, Godang, K, Mollerup, C, Fougner, S.L, Pernow, Y, Aas, T, Hessman, O, Rosén, T, Nordenström, J, Jansson, S, Heck, A[†]. and Bollerslev, J[†].

Journal of Bone and Mineral Research. 2021; 36: 3-11.

DOI: 10.1002/jbmr.4199h

Paper II: Mortality and Morbidity in Mild Primary Hyperparathyroidism: Results From a 10-Year Prospective Randomized Controlled Trial of Parathyroidectomy Versus Observation

Pretorius M[*], Lundstam K[*], Heck A, Fagerland MW, Godang K, Mollerup C, Fougner SL, Pernow Y, Aas T, Hessman O, Rosén T, Nordenström J, Jansson S, Hellström M [†], Bollerslev J[†].

Annals of Internal Medicine. 2022; Jun; 175(6):812-819

DOI: 10.7326/M21-4416

Paper III: Positive effect of parathyroidectomy compared to observation on BMD in a randomized controlled trial of mild primary hyperparathyroidism

Lundstam, K[*], Pretorius, M[*], Bollerslev, J, Godang, K, Fagerland, M, Mollerup, C, Fougner, SL, Pernow, Y, Aas, T, Hessman, O, Rosén, T, Nordenström, J, Jansson, S, Hellström, M[†], Heck, A[†].

DOI: 10.1002/jbmr.4763

[*] Shared first authorship

[†] Shared last authorship

Summary:

What is the problem and what is known about it so far?

In the neck, close to the thyroid gland, there are four much smaller parathyroid glands. Their main function is to control the level of calcium in the blood. When calcium is too low, the glands produce parathyroid hormone (PTH) that moves calcium from bones into the blood. These glands can develop a benign tumor that produces too much hormone, which makes calcium levels too high. This disease, called primary hyperparathyroidism, is a common finding in females after the menopause. Minor increases in calcium may not cause symptoms, but larger increases can cause nonspecific symptoms, such as muscle weakness, fatigue, bone pain, and specific symptoms, like nausea, confusion, and increased thirst. When large increases last a long time, they can lead to fractures, kidney and heart problems and early death.

Why did our research group do this particular study?

Surgery to remove the tumor is the only treatment known to prevent these negative outcomes. However, it is uncertain whether surgery prevents these outcomes in patients with only marginal increases in calcium levels.

Who was studied?

One hundred and ninety-one patients in Sweden, Norway, and Denmark with benign parathyroid tumors and minor increases in calcium levels who were otherwise healthy.

How was the study done?

Half of the patients were randomly assigned to have surgery, and the other half to observation without surgical treatment. The primary end-point of the study was mortality. Further, secondary end-points were fractures, cardiovascular disease, malignancies, kidney disease and quality of life.
Both groups of patients were monitored for at least 10 years. During this time, they received appropriate treatment, including surgery to remove tumors if it later became necessary.

What did our research group find?

About equal numbers of deaths, fractures, cardiovascular disease and malignancies occurred in the groups. In both groups, fewer patients died than anticipated. The bone mineral density decreased significantly in the observation group, however this did not lead to increasing numbers of fractures in this group during the observation period. Some subtle improvements in quality of life were found in the operated group, but with uncertain clinical relevance.

What were the limitations of the study?

We were unable to recruit as many patients to the study, as planned. Nevertheless, this is the largest study of its kind and, so far, with the longest follow-up period.

What are the implications of the study?

Early surgery does not appear to reduce negative outcomes in otherwise healthy patients with benign parathyroid tumors and slightly elevated calcium levels. There are long-term benefits in bone mineral density with surgery but otherwise no obvious clinical improvements after ten years of follow-up. The study is an important contribution to guide both patients and clinicians in shared decision-making regarding operation for this common disease.

Sammendrag:

Hva vet vi om problemstillingen hittil?

Rett bak skjoldbruskkjertelen ligger fire små kjertler kalt biskjoldbruskkjertler. Deres hovedfunksjon er å opprettholde et stabilt nivå av kalsium i blodet. Hvis kalsiumnivået blir lavt skilles PTH (parathyreoideahormon) ut fra biskjoldbruskkjertlene, Dette hormonet flytter kalsium fra skjelettet til blodbanen. Biskjoldbruskkjertlene kan utvikle godartede svulster som produserer for mye hormon og fører til et for høyt kalsiumnivå. Hyppigst ser man denne tilstanden, kalt primær hyperparathyreoidisme, hos kvinner etter overgangsalderen, opp til 2-3 % i denne gruppen. En liten økning i kalsiumnivået fører vanligvis ikke til symptomer, men ved mer uttalt økning i kalsiumnivået kan man ha uspesifikke symptomer som muskelsvakhet, trøtthet, skjelettsmerter eller mer spesifikke symptomer som kvalme, forvirring, og økt tørste. Hvis høye verdier vedvarer over lang tid kan det føre til benbrudd, nyrestein, hjerteproblemer, og tidlig død.

Hva har vår forskningsgruppe gjort i denne studien?

Fjerning av en overaktiv biskjoldbruskkjertel er den eneste virksomme behandlingen ved høye verdier og for å hindre komplikasjoner. Det er usikkert hva man skal gjøre hvis det kun er lett forhøyede verdier av kalsium. Skal man da operere, eller kan man se det an siden tilstanden oftest holder seg stabil over lang tid?

Hvem ble med i studien?

191 pasienter fra Sverige, Norge og Danmark uten vesentlige andre sykdommer, men med mild primær hyperparatyreoidisme og lett forhøyet kalsiumnivå.

Hvordan ble studien gjennomført?

Ved randomisering ble det bestemt om pasienten ble operert eller ikke. Primært endepunkt i studien var mortalitet. Predefinerte sekundære endepunkter var frakturer/ bentetthet, kardiovaskulær sykdom, malignitet, nyresykdom og livskvalitet.
Begge grupper ble så fulgt i minst 10 år. Hvis en pasient i observasjonsgruppen utviklet komplikasjoner ble de operert.

Hva fant vår forskningsgruppe?

I begge grupper døde omtrent like mange pasienter. Omtrent samme antall utviklet andre komplikasjoner som benbrudd, kardiovaskulær sykdom og kreftsykdom. Bentettheten ble signifikant redusert i observasjonsgruppen, men dette ledet ikke til

færre brudd hos disse pasientene. Små forbedringer ble funnet i gruppen som ble operert men med usikker klinisk relevans.

Hvilke begrensninger har studien?

Vi klarte ikke å få så mange pasienter i studien som opprinnelig planlagt. Likevel er denne studien den største og lengste studien knyttet til denne problemstillingen.

Hva betyr denne studien for pasientene?

Fjerning av biskjoldbruskkjertelen ved kun lett forhøyede kalsiumverdier og ukomplisert mild primær hyperparatyreoidisme ser ikke ut til å redusere dødeligheten eller forekomst av brudd, kardiovaskulær sykdom eller cancer i et 10 års perspektiv. Operasjon viste seg fordelaktig med tanke på bentetthet men forøvrig uten åpenbare kliniske forbedringer etter 10 års oppfølging.
Studien er et viktig bidrag for å kunne gi bedre råd til pasienter og støtte samvalg med pasientene rundt operasjon av biskjoldbruskkjertelen.

Introduction/ Background

Physiology of Parathyroid Hormone

Parathyroid hormone is essential for the tightly regulated calcium and phosphate homeostasis (1). It also facilitates the activation of Vitamin D through the stimulation of 1-alfa hydroxylase in the proximal convoluted tubule of the kidney (see *Figure 1a:*). Through a feedback mechanism, PTH normally declines when calcium is rising (2). Because of a rapid synthesis and storage in the parathyroid gland, PTH can rise within seconds after a low calcium level is detected. PTH has a short half-life (of a few minutes), which can be seen in the intraoperative PTH measurements done when removing a parathyroid adenoma (3).

In the skeleton, PTH increases bone resorption through receptor activator nuclear factor kappa B ligand (RANKL) thereby stimulating the cells responsible for breaking down bone tissue (the osteoclasts). PTH stimulates osteoclast activity, differentiation, and survival. PTH can also stimulate bone formation both through direct stimuli of the bone-forming cells (osteoblasts) via the PTH1-receptor and through inhibition of bone inhibitors such as sclerostin (4).

In the kidney, PTH stimulates the reabsorption of calcium mainly through calcium ion channels such as the transient receptor potential vanilloid 5 (TRPV5) (5). PTH regulates the kidney dependent conversion of 25-OH vitamin D to its biologically more active form 1.25-dihydroxy vitamin D (6). Also of importance, PTH downregulates the two sodium-dependent phosphate cotransporter, NaPi IIa and NaPiIIc, leading to increased phosphaturia (7).

Although much is known about the metabolic process occurring within bones and the kidney there are still uncertainties on how PTH affects other target tissues such as skeletal and cardiac muscles, central nervous system (CNS), and adipocytes (8). PTH works through the specific PTHR1 and PTHR2-receptors (9, 10). Both are expressed widely throughout the body. There are clues from basal research that PTH can affect skeletal muscle, cardiac muscle and energy metabolism (11, 12). However, the extent and clinical implication in humans needs further understanding. The widely distributed PTH-receptors in the areas in the brain that regulate fear, anxiety, pituitary hormones and nociception suggests a functional rational for effect in the CNS (13). However, the relevance and implications of these findings needs further study.

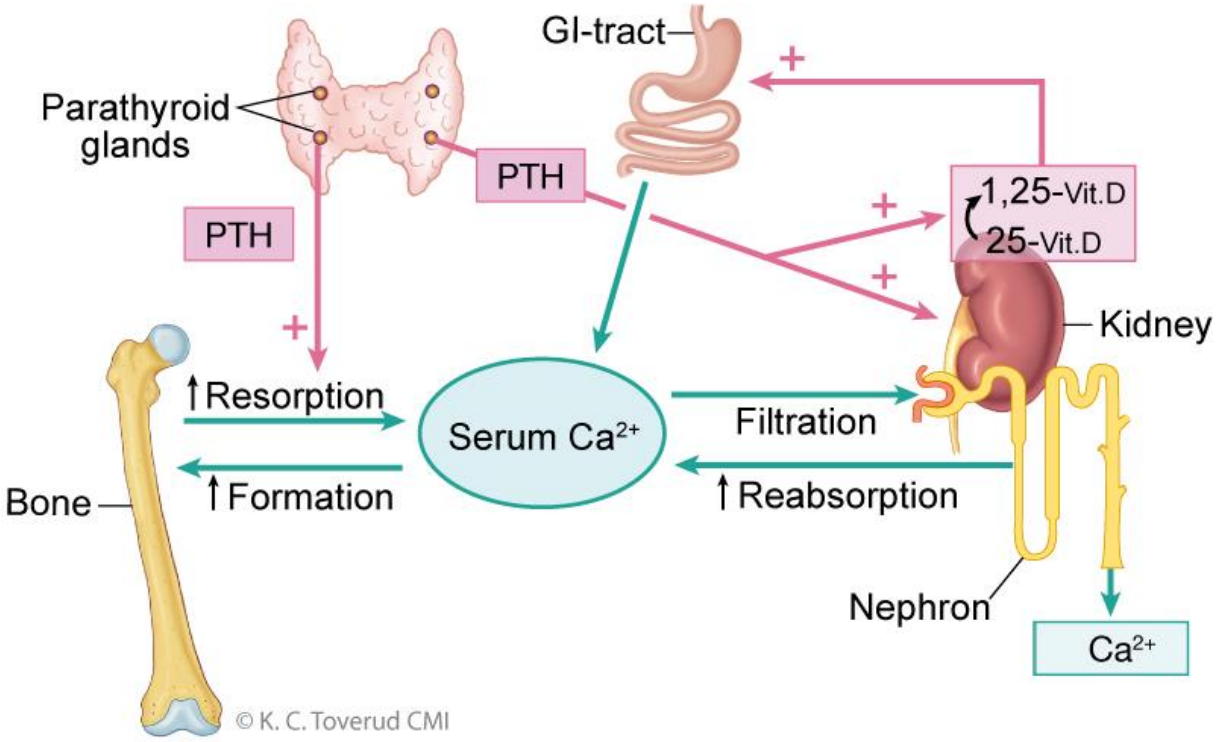

Figure 1a: PTH-centered illustration of the effect of PTH on calcium homeostasis. Figure made on commission by Kari C. Toverud for the SIPH-study group and published in Best Practice Res Clinical Endocrinology and Metabolism (14). Printed with permission.

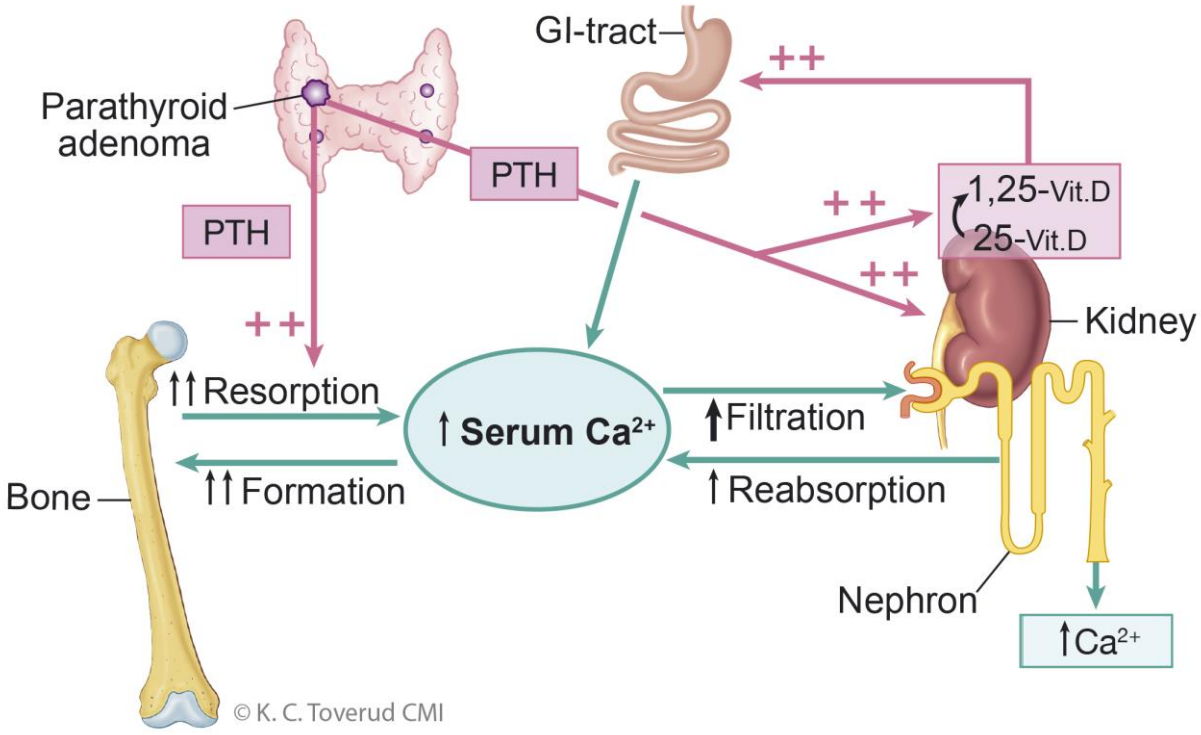

Figure 1b: Calcium homeostasis in Primary Hyperparathyroidism. Figure made on commission by Kari C. Toverud and published in EJE as part of the PARAT-group (15). Figure printed with permission.

The History of Parathyroid Hormone

The knowledge of the parathyroid hormone has increased significantly during the last two centuries. Advances in anatomy, physiology, medicine, surgery, biochemistry, and genetics have all contributed to our current understanding. Still, there are numerous questions unanswered on both diagnosis and treatment regarding many parathyroid-related diseases. The field of endocrinology is a young field starting as a clinical science in the nineteenth century. The parathyroid glands were discovered by Sir Richard Owen in 1852 (16). He did a necropsy of a great Indian Rhinoceros and described these yellow glands, "the size of small peas", attached to the thyroid gland. Since this discovery was published in a zoological paper it went unnoticed for many years until the Swedish pathologist Sandstrøm got to name the glands ("glandulae parathyroidae") after first discovering them in dogs (17). He went on to identify the glands in rabbits, cats, ox, and horses. After this, he did autopsies on humans and found the same small glands close to the thyroid. His work was rejected by journals in Germany and he ended up publishing in a local Swedish journal in 1880 (17).

Many thought the findings of Owen, Sandtrøm, and others were fetal remnants. It was not before a French physiologist (Eugene Gley) in 1891 observed that tetany and death occurred in dogs when thyroid removal included the parathyroid glands that linkage to disease was made (18). A few years later (1898) Meussu managed to treat tetany following parathyroidectomy in dogs by injecting extracts from horse parathyroids, thereby confirming that a substance from the glands was indeed the cause of this disease. In the coming decades, this technique was refined and one of the hallmark studies on this treatment was written by James B. Collip in 1925 (19). Several theories were debated in the early 20th century but through a series of experiments, Collip showed that acid extracts of the parathyroid gland completely relieved the tetany that followed parathyroidectomy. He thereby established the parathyroids as an endocrine gland, which secreted hormone (parathyroid hormone, PTH). James Collip is often cited as the one who discovered parathyroid hormone and was able to isolate an extract of it to treat tetany. He has since been even more renowned for his work with Banting and Best in purifying insulin, but his pioneer work with PTH is equally important in the parathyroid field (20).

Figure 2: *James Bertram Collip- 1892–1965 Portrait © National Portrait Gallery, London, printed with permission*

In approximately the same period, there was a growing understanding of the relationship between bone disease and the hyper-functioning parathyroid glands. Some

earlier work had eluded to this connection but it is first with the work of von Recklinghausen in 1891 that this link was acknowledged (21). For many years, primary hyperparathyroidism (PHPT) was known as von Recklinghausen`s disease of bone. The clinical manifestations of severe parathyroid disease (the classical symptoms of "stones, bones, groans and moans") were of course present centuries before the anatomical and physiological understanding of the parathyroid gland. The existence of osteitis fibrosa cystica (pathognomonic to hyperparathyroidism) had been found in patients dating back to ancient Egypt (22). These findings are probably more due to the well-preserved skeletal specimens from this culture and there are no reasons to believe that the disease was unique to this North-African area. Davis Colley presented a clinical history and autopsy findings in 1884 of a young girl who died of a combination of skeletal and urinary tract disease (23). Retrospectively the histological findings were probably due to hyperparathyroidism but were at the time described as osteomalacia.

The first physician to describe an enlarged parathyroid gland in relation to osteitis fibrosa cystica was Schlagenhafer. He has also been accredited the concept that the bones were the result of another disease rather than the cause of the disease. This concept made treatment possible and in 1925 the first patient was operated on by Felix Mandl (in Vienna) with the removal of a parathyroid adenoma. It was described as dramatic improvement and total healing of symptoms but the patient died seven years later from recurrence of the disease (24). One of the great physicians of all time, Fuller Albright, summed up knowledge about both pathophysiology and histology and managed to categorize the disease in the book "the parathyroid glands and metabolic bone disease" published in 1948. His categorization of primary and secondary hyperparathyroidism made decades before laboratory tests could support this distinction is still used today (25). Several breakthroughs in both molecular biology, cell biology, genetics, and laboratory medicine made it possible to both determine the PTH structure and synthesize the biological active segment of PTH (26). This was followed by the successful cloning of the PTH-receptor (27) which could now give rise to genetic explanations of many of the diseases described clinically by Albright in his book from 1948. The synthesis of PTH meant that previously serious and mortal diseases caused by hypoparathyroidism could now be treated with injection therapies.

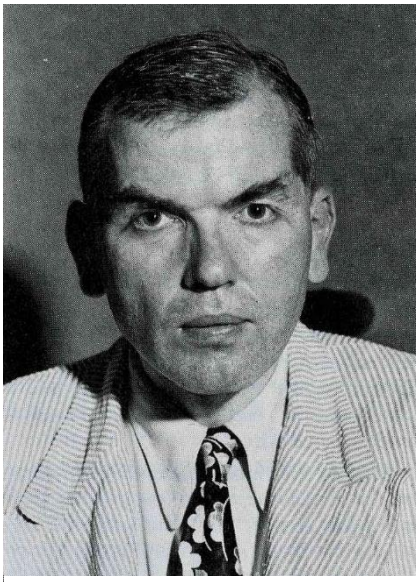

Figure 3: *Fuller Albright 1900-1960. Picture published in the 1970-edition of the Harvard Medical Alumni Bulletin without copyright. Permitted use as public domain.*

Up to the mid-20th century, mainly symptoms had been relevant for the diagnosis of parathyroid disease. Even though mineral ions were possible to measure in the early decades of the 20th century, the methods were tedious and the results were often inexact. In the late 20th century, the disease spectrum changed, with many patients now diagnosed without any of the classical symptoms nor with any other symptoms. This was described as asymptomatic hyperparathyroidism. The incidence of primary hyperparathyroidism has increased dramatically over the last four decades. This increase is believed to be both due to an awareness of the disease but also attributed to the readily available calcium measurement as a routine blood analysis and the availability of PTH- immunoassay methods. How to treat and follow these patients without symptoms led to a challenge in the medical community, which still exists today (15). During the short history of PHPT the change from rare disease with multi-organ affection to today's mild disease without specific symptoms, is the basis for this thesis, as the optimal management of mild PHPT is still largely unknown.

Primary Hyperparathyroidism

Today, PHPT is a common endocrine disorder, especially in peri- and postmenopausal women (28, 29). The incidence and prevalence vary globally presumably related to the availability and screening practice of calcium and PTH-measurements (29). Recent estimates suggest a prevalence in the United States of 233 per 100.000 women and 85 per 100.000 men with the highest overall prevalence in women aged 70-79 years (30). Data from Scandinavian countries report a prevalence of 2-5% in peri-and postmenopausal women (28, 31).

PHPT is characterized by elevated calcium levels and inappropriately high levels of PTH (32). In countries with a highly developed health care system, the mild (previously termed asymptomatic) form is now predominant. In regions where routine doctor visits without symptoms are rare, the symptomatic disease, often with organ manifestations, is still most common (33). Case reports that describe serious disease with multiple symptoms and devastating organ involvement are still published from developing countries (34, 35).

PHPT is caused by excessive secretion from one (approximately 85%) or multiple of the four parathyroid glands (36). In overt cases, this can lead to increased bone turnover, secondary osteoporosis, and kidney diseases such as nephrolithiasis and chronic kidney disease (CKD) (34). Removing the hyper-functioning gland by parathyroidectomy (PTX) is the only curative treatment. PTX has a high success rate. In its most serious form, PHPT can result in disequilibrium hypercalcemia (see below) and hypercalcemic crisis with multi organ complications and fatal outcome (37).

However, severe hypercalcemia is rare, and most commonly the diagnosis of PHPT is made in patients without classical symptoms and severe complications.

Several studies have reported remarkably stable calcium values in mild PHPT over decades (38, 39). This could be attributed to the concept of equilibrium hypercalcemia due to a set point error described by Parfitt et al. already in 1979 (40). Parfitt characterized equilibrium hypercalcemia as calcium above the normal reference value maintained at the same level with few fluctuations for years. The hypothesis was that this pattern seen in mild to moderate PHPT is caused by an increase in calcium bone resorption and balanced by an increase in renal tubular excretion. The model of equilibrium hypercalcemia is illustrated in Figure 4. The seemingly non-progressive hypercalcemia in PHPT in most patients with mild disease is why PTX in these patients is still widely debated.

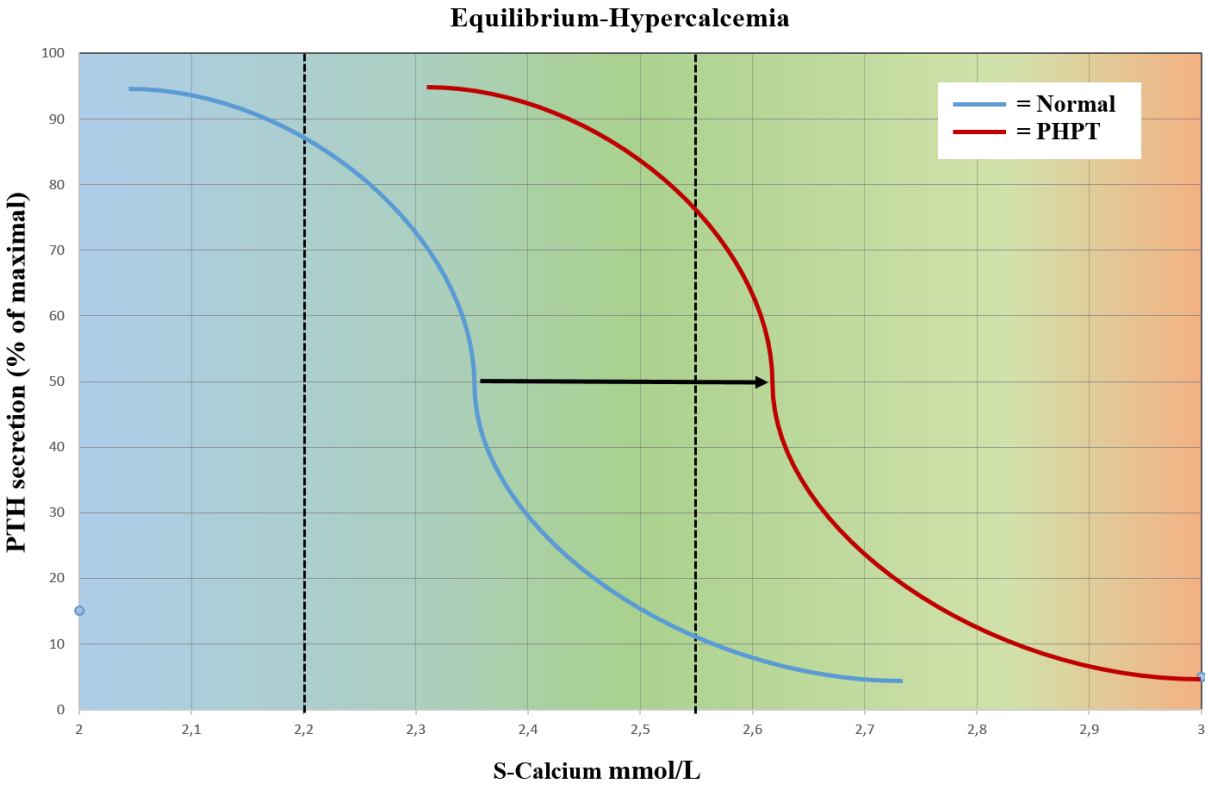

Figure 4: Model of Equilibrium Hypercalcemia in both normal physiology and PHPT. Dotted lines = upper and lower reference values for s-calcium in mmol/l.

Since the start of the Scandinavian Investigation of Primary Hyperparathyroidism (SIPH-study) in 1998, the definition of mild disease has changed several times. The term asymptomatic PHPT was previously used to separate these patients from patients with symptoms of (often severe) hypercalcemia. During the nineties the term mild PHPT was established since many of these symptoms were unspecific and common in people without hypercalcemia. The last consensus report now separates the

asymptomptomatic patience into two forms; with and without target organ involvement (41). Until then, there has been no universal definition of mild disease. This has hampered many meta-analyses and long-term studies, which often include both mild and severe forms of PHPT.

International consensus guidelines on the management of asymptomatic PHPT have been issued five times since the first one in 1991 with the last guideline recently published in 2022 (41-45). European consensus statements have also been evolved mainly through the PARAT-program initiated by the European Society of Endocrinology (46). Generally, the indications for PTX have expanded through the years.

Diagnosis of primary hyperparathyroidism.

Since most patients in countries within a resource-rich health care system are incidentally identified without specific symptoms or clinical findings, the diagnosis is largely based on routine serum calcium measurements. Still, case finding in patients with kidney stones or osteoporotic fractures can lead to the diagnosis. The combination of elevated serum calcium levels and high/inappropriate high PTH-levels is the key to the diagnosis (41, 47). Several differential diagnoses should be considered before concluding that the patient has PHPT. Drugs affecting calcium handling and hereditary history are important topics to elaborate upon before diagnosis (see differential diagnosis below).

There is an ongoing discussion on what is the best measurement in regard to calcium. Since the introduction of serum auto-analyzers, the standard has been to recommend serum albumin-corrected calcium in the diagnosis of PHPT. There are several ways of albumin correction, the most commonly used is: Albumin-adjusted calcium (mmol/L) = total calcium (mmol/L) + 0.02 [40 – albumin (g/L)] (48). This has also been used as basis for the diagnosis in the SIPH-study.

Differential diagnosis

There are several important differential diagnoses to consider before deciding on treatment options for the patient. When both calcium and PTH are elevated (or inappropriately high), both drugs and hereditary conditions are possibilities. Thiazides are still the most commonly used antihypertensive agents worldwide (49). Thiazides lead to hypercalcemia probably because of a combination of inhibition of TRPV5 and calcium reabsorption due to volume depletion (50, 51). Although none of these mechanisms directly lead to PTH secretion, PTH will in many cases be inappropriately high related to the calcium level (51, 52).

Lithium is another drug that could change the management of hypercalcemic hyperparathyroidism. Lithium interacts with the calcium-sensing receptor (CASR) and leads to a set point elevation of calcium concentration (53). After prolonged use, this can lead to the development of hyperplasia or adenomas of the parathyroid gland. In many of these patients, hypercalcemia will resolve when Lithium is discontinued, but not for all (54). The incidence of multiglandular disease is higher in these patients which could lead to a different surgical approach (55).

Several genetic diseases are associated with hyperparathyroidism (29). The most common are familial hypocalciuric hypercalcemia (FHH), multiple endocrine neoplasia type 1 and 2a, and mutations in the CDC73 gene.

Parathyroid cancer is another rare, however important differential diagnosis. Generally, these patients have a higher calcium level and larger tumors. Since most patients do not undergo routine imaging procedure this latter distinction is not always clinically recognized (56). There are no specific clinical, biochemical, or radiological characteristics that can distinguish malignant from benign PHPT (57). Thus, the diagnosis of malignancy is often first made after surgery.

Mortality in Primary hyperparathyroidism.

Prior to the SIPH-study little was known about mortality in mild PHPT. Studies had demonstrated an increased mortality in both hospitalized and non-hospitalized patients with PHPT (58, 59). During the course of our study, several epidemiological and observational studies demonstrated this increased risk of mortality, cardiovascular disease, and malignancy in PHPT (60-63). This was supported by studies with indirect findings of increased artery stiffness, dyslipidemia, and ventricular hypertrophy (64-66). However, results from other non-randomized trials did not support these findings (67-69). So far, only two other randomized controlled trials have compared PTX with observation in mild PHPT. These had one and two years of follow-up, respectively, and both were limited by small sample size and absence of long-term follow-up (70, 71). A recent randomized controlled trial (RCT) by Ejlsmark-Svensson et al. indicated a decreased risk of cardiovascular disease by lowering cholesterol levels after PTX (72). However, the study had a short duration (3 months) and included patients with both mild, moderate, and severe hypercalcemia. In the SIPH study, the metabolic and cardiovascular risk factors (including cholesterol) were not improved by PTX after 5 years (69).

Due to short duration, retrospective and non-randomized design, the results of the previous studies cannot be generalized (external validity) and may be confounded by uncontrolled factors. Randomized controlled trials on hard end-points such as

mortality, cardiovascular disease, and malignancy are missing on this common endocrine disease.

To our knowledge, the SIPH-study is the first randomized controlled trial designed to prospectively compare long-term survival and key morbidities such as cardiovascular, cerebrovascular and renal disease as well as malignancy, between observation and parathyroidectomy.

Quality of Life in Primary Hyperparathyroidism

Quality of Life (QoL) is defined (WHO) as “an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns” (73). During the last decades, the centrality of the patient's point of view in monitoring medical outcomes has become essential in clinical treatment and research. An ongoing discussion is how PHPT affects QoL. There is both anatomical and clinical reasoning behind the assumption that also mild PHPT could have cognitive and psychopathological effects. Both Calcium Sensing Receptor (CASR) and PTH1 and PTH2-receptors are found abundantly in the brain (74). Especially for excitable cells such as neurons, the tight control of calcium homeostasis is crucial. Calcium is a small ion that easily cross membranes and, in the CNS, it is responsible for neurotransmission/excitability (74).

In 2015, a meta-analysis on the randomized controlled trials on this subject suggested that PTX might be associated with better QoL in asymptomatic PHPT (75). Only three studies were included (70, 71, 76) with effect measurements at 6, 12, and 24 months. Previous results from the SIPH-study contributed to a majority of the weight in this meta-analysis at both 12 and 24 months (76). An even more recent systematic review included seven observational single-arm studies and four randomized controlled trials (77). Almost all the included studies showed some statistically significant differences, favoring PTX.

Management of Primary Hyperparathyroidism

Surgical management

International guidelines for the surgical management of asymptomatic PHPT were published for the first time in 1991 (42). For patients that meet the surgical criteria, parathyroidectomy is generally advised (78). In the current guidelines of 2022 patients with symptomatic disease, as well as patients with renal or bone manifestations or age below 50 years should all be advised surgery. In addition, patients with calcium levels

0.25 mmol/l above the upper limit of normal as well as increased urinary calcium excretion (above 250 mg/day) are recommended PTX (41).

Since the first described parathyroidectomy by Felix Mandl in 1925 numerous advances has been made in the surgical treatment of PHPT. The evolvement of localization techniques such as Sestamibi scintigraphy, high-resolution neck ultrasound and high resolution multiphase contrast medium-enhanced computed tomography (4D-CT) has led to better preoperative imaging and surgical planning. Therefore, the majority of PTX today are performed as minimally invasive surgery with selective, targeted removal of a single gland (79). In many centers, intraoperative PTH serum measurements are done to assess successful operation, or to decide if further exploratory surgery is needed (80). The most recent guideline on surgical management of PHPT states that intraoperative PTH-measurement in routine use is not essential in patients with two concordant preoperative localization imaging studies (79). However, it can lead to a marginal increase in cure rate but with longer operation time and potentially unnecessary contralateral explorations (81). As surgery has a high cure rate and low frequency of complications there has been a tendency to widen the indications for this treatment during the last decades.

Medical treatment:

Medical (non-surgical) treatment is an alternative for patients that meet the surgical criteria but cannot go through PTX. This could be because of contraindications to surgery, unwillingness, comorbidities, or if previous surgery has been unsuccessful. Most of these medical alternatives and the evidence of their efficacy in PHPT have become available during the last two decades, and was therefore not mentioned in the original SIPH-protocol. However, none of the medical treatments available today will cure the disease (82).

Bisphosphonates

Bisphosphonates are a class of medication that inhibit bone resorption by reducing osteoclast function (83). Several bisphosphonates have been used to treat PHPT. Leere et al. published a systematic review of the efficacy related to the different kinds of bisphosphonates (84). Seven studies reported on Pamidronate, all with a short follow-up and no bone mineral density (BMD) data. There was a minute decrease in serum calcium (-0.034 mmol/L) but five of the studies showed an increase in PTH. Alendronate was studied in twelve of the included studies of the systematic review. The decrease in serum calcium was also marginal (-0,05mmol/L). All the long-term studies reported an increase in BMD in the spine and hip with stable values in the distal radius. All studies reporting on bone turnover markers demonstrated a rapid and marked decline in these biomarkers.

During the last decade, treatment with intravenous bisphosphonates (Zoledronic acid) has come available. With its long-time effect (years) the possibility of combining PTX and Zoledronic acid has been a treatment alternative evaluated both in observational and controlled trials (85-87). The minute and possibly transient effect on serum calcium makes bisphosphonates a viable option for skeletal protection but would rarely affect symptoms or other organ manifestations related to hypercalcemia in PHPT (88). The studies above were all performed after initiation of the SIPH-study and therefore this treatment alternative is not included in the SIPH study protocol.

Cinacalcet

Cinacalcet is a calcimimetic agent approved for treatment of primary and secondary hyperparathyroidism in the European Union in 2004. Cinacalcet increases the sensitivity of the calcium-sensing receptor (CASR) thereby decreasing the synthesis and secretion of PTH (89). Cinacalcet reduces the set point of the PTH- calcium axis as illustrated in Figure 4 (90). A recent systematic review and meta-regression included 28 studies on the use of Cinacalcet in PHPT (91). Serum calcium and PTH levels were significantly reduced with a normalization rate of 90% across studies. None of the studies had a positive effect on bone mineral density and can therefore not be used for skeletal protection. The combination of bisphosphonate therapy and Cinacalcet has been studied and could be a viable alternative for management where PTX is not an alternative in osteoporotic patients with symptoms of hypercalcemia (92).

Denosumab

Denosumab is a human monoclonal antibody that blocks the binding of RANKL, thereby inhibiting osteoclast development and activity (93). It has mainly been used to treat postmenopausal osteoporosis but the protective effect on bone in PHPT has recently been demonstrated in a randomized double-blind placebo-controlled trial (94). It showed a significant increase in BMD in the lumbar spine, femoral neck, and total hip after one year compared to placebo. Other, non-randomized trials have demonstrated a similar effect and Denosumab seems to provide skeletal protection in the same manner as in the treatment of postmenopausal osteoporosis (95, 96). Similar to bisphosphonates, Denosumab lowers bone turnover markers but does not reduce calcium or PTH-levels (95).

Aims of the Study:

Mild Primary Hyperparathyroidism is now the predominant form of hyperparathyroidism, probably due to increased routine blood sampling including measurements of both calcium and PTH. The present thesis focuses on the predefined primary and secondary endpoints in the SIPH protocol from 1998. The aim of this study was to investigate if PTX has an effect on long-term mortality and morbidity in mild PHPT.

The **primary end-points** as defined in the original protocol were registered-based mortality and cause of death.

The predefined **secondary end-points** were:

- Cardiovascular disease
- Malignant disease
- Skeletal disorders including clinical fractures and bone densitometry outcome
- Kidney disease including both kidney function and stone formation in the urinary tract.
- Quality of life

The aims of this thesis were focused on end-of study data on all the predefined aims:
Mortality (primary aim) was described in Paper II.
The 10-year data for the secondary aims were described in:
→ Paper I (quality of life)
→ Paper II (cardiovascular disease, malignancy, clinical fractures and kidney stones)
→ Paper III (bone densitometry and clinical fractures)

Materials and Methods

Study design and patients

The SIPH-study (Scandinavian Investigation of Primary Hyperparathyroidism) was an investigator-initiated, prospective, randomized controlled clinical trial. The final protocol was released on March 1, 1998 and the first patient included October 1, 1998. After 6.74 years of inclusion (the last patient included June 29, 2005) there was a preplanned observation period of 10 years. A total of 191 patients were included; 96 were randomized to observation and 95 to parathyroidectomy. The patients were randomized by a concealed envelope system developed in collaboration with Medstat Research. Block randomization was applied, with block sizes randomly set between two and ten. The eight Scandinavian study centers received 100 randomization envelopes each. This procedure assured a balance between and within the study centers. No further stratification factors were added. After study cessation, we have published the entire randomization procedure; https://www.acpjournals.org/doi/suppl/10.7326/M21-4416/suppl_file/M21-4416_Supplement.pdf.

Inclusion criteria were:

- Primary hyperparathyroidism
- Untreated mild hypercalcemia (defined between 2.60 mmol/l to 2.80 mmol/l or ionized calcium between 1.40 and 1.51 mmol/l)
- Informed consent

The eight exclusion criteria were: age below 50 or above 80, former neck surgery, osteitis fibrosa cystica, kidney failure, psychiatric disease, kidney stones, familiar forms of PHPT, and other serious medical illnesses.

Study Schedule and procedures:

During the 10 years of follow-up, patients were scheduled for annual visits. The complete investigation scheduled for each visit is illustrated in Table 1.

Examinations	Inclusion	3 mo	1y	2y	3y	4y	5y	6y	7y	8y	9y	10y
Medical history	X	X	X	X	X	X	X	X	X	X	X	X
Medications	X			X			X					X
Cardiac disease	X			X			X					X
Cerebrovasc disease	X			X			X					X
Malignancy	X			X			X					X
Peripheral fractures	X			X			X					X
Objective analyses	X	X	X	X	X	X	X	X	X	X	X	
Blood pressure	X	X	X	X	X	X	X	X	X	X	X	X
ECG	X			X			X					X
Patho-anatomy *(operated individuals)*												
Histology	X											
Resected tissue weight	X											
Biochemistry												
Hb	X	X	X	X	X	X	X	X	X	X	X	X
S-Na (sodium)	X											
S-K (potassium)	X											
S-Calcium	X3	X	X	X	X	X	X	X	X	X	X	X
S-Albumin	X3	X	X	X	X	X	X	X	X	X	X	X
S-Phosphate	X											
S-PTH (1-84)	X2	X	X	X	X	X	X	X	X	X	X	X
S-B-phosphatase	X											
S-Creatinine	X	X	X	X	X	X	X	X	X	X	X	X
dU-Creatinine	X			X			X					X
20 ml f-serum	X			X			X					X
5 ml whole blood	X			X			X					X
20 ml d-Urine	X			X			X					X
Radiological exams												
Kidneys and urinary tract (KUB)	X			X			X					X
Lumbar spine	X			X			X					X
Osteodensitometry												
Whole body	X		X	X			X					X
Forearm (dominant)	X		X	X			X					X
Spine (ap)	X		X	X			X					X
Hip (right)	X		X	X			X					X
Quality of life (questionnaire)	X			X			X					X

Table 1: Investigation schedule throughout the SIPH study. Picture from the original Protocol of 1998.

Approvals and informed consent

The study was approved by the regional ethical committees in all the three countries: S-96107 and 2017/2413-REK Sør-Øst (Norway), DNR-96349 and 18721/2019

(Sweden) and S-96007 (Denmark). All patients signed informed consent in their language. The study was registered at ClinicalTrials.gov: NCT00522028

Quality of Life

For the QoL assessment two main health status instruments were used, the Short Form 36 (SF-36) Version 1 and the Comprehensive Psychopathological Rating Scale (CPRS).

SF-36 is a generic quality of life measurement. It is a condensed version of the questionnaire used in the Health Insurance Experiment (97). Several versions were attempted during the Medical Outcome Study (MOS) (98). The continuous shortening of these questionnaires was a response to study participants refusing to complete time-consuming surveys (99). The SF-36 is a questionnaire consisting of 35 questions giving rise to eight dimensions of physical (four dimensions) and mental (four dimensions) health (Figure 5). The last question is about self-reported health improvement during the last year. In the SIPH-study, SF-36 was used as a self-administrated questionnaire completed at the time of the study visit. SF-36 could also be done as a telephone interview, mail-back questionnaire, or face-to-face interview. In these settings, special considerations are advised, including written patient information when using mail-back or scripts for telephone interviewing or face-to-face interviews (100). Even though SF-36 is often termed a generic survey, some content had to be removed when shortening the form. Most noteworthy is the exclusion of cognitive functioning, sleep, family functioning, and sexual functioning, none of which are covered by the SF-36 (99). By today's standard SF-36 is not regarded as short, and even more comprised versions (like the SF-12) have been made available after the initiation of the SIPH study (101). In the 1990's efforts were made by the IQOLA project to translate and validate SF-36 to different languages thereby making them validated and available in the three Scandinavian languages in time for the SIPH-study (102). This was done in a standard protocol including multiple forward and backward translations as well as qualitative evaluations of the translations and scaling assumptions (103). In this process, not only direct interpretations were made, but also changes in meaning to accommodate local understanding. This could be exemplified by the American translation of walking several blocs (Q3h), to walking half a mile (UK version) to walking a few hundred meters (Swedish version).

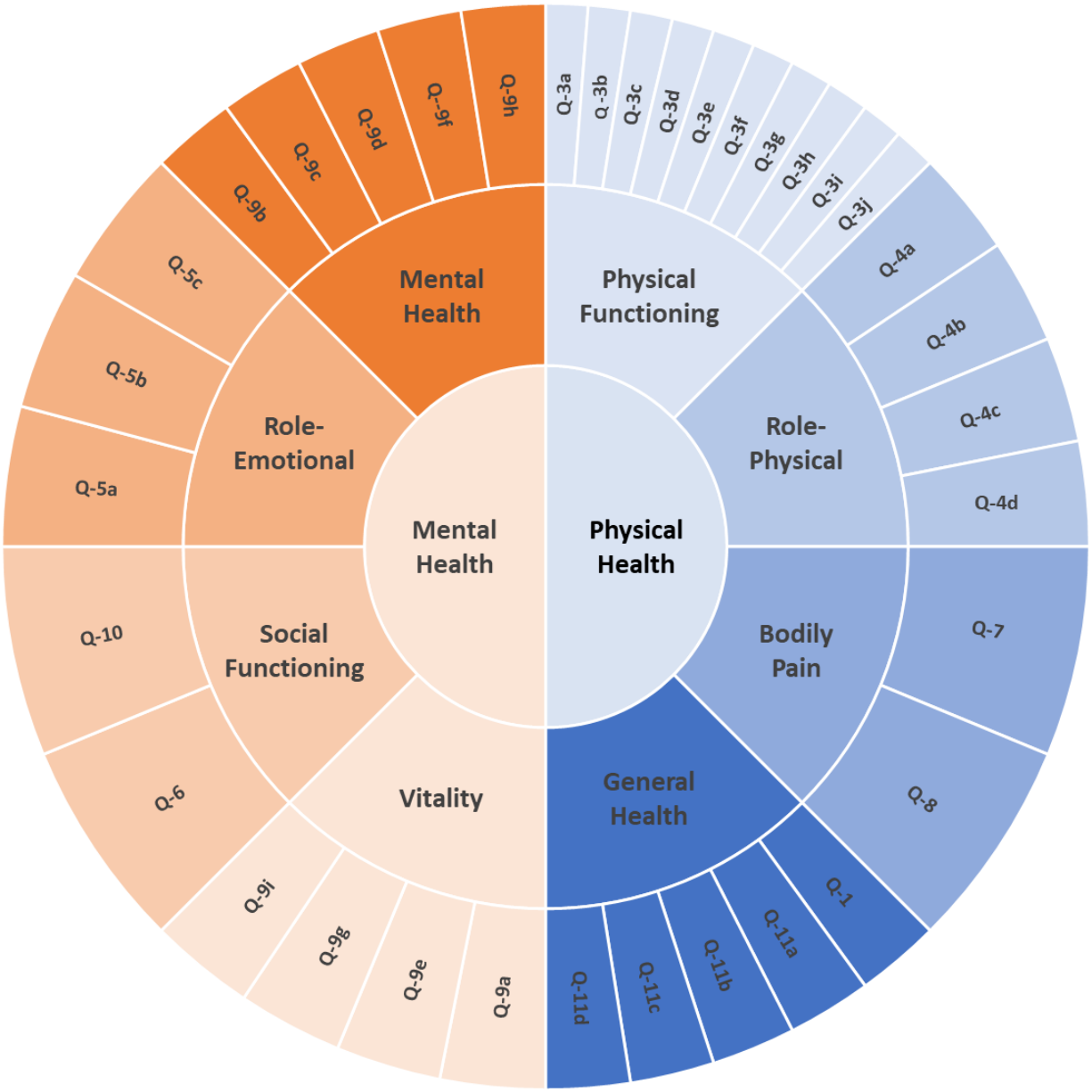

Figure 5: Structure and contextual overview of the SF-36-V1. Copyright Pretorius, M

As stated in Table 1, the patients filled in the forms at baseline, two, five, and ten years. At baseline, 179 had validated SF-36 forms. Of the 129 patients completing the trial, 122 had validated SF-36 questionnaires available at the 10-year visit. Since these data were gathered and initially registered over a period of twenty years three SIPH authors went through each answer cross-checking the original CRF with the database in 2018. This gave rise to 24,137 individual patient answers with 631 (2.6 %) missing values.

The next step involved transforming the raw dimension score for each patient to a 0 to 100 scale using the formula: Transformed dimension score = ((Actual dimension score

– lowest possible dimension score) / possible raw score range) x 100. According to the SF-36 manual (104), a dimension scale score could be calculated if above 50 % of the answers in this dimension are valid leading to a total number of missing values of 2.2%. Although possible, no imputation or other forms of filling in the missing values was used. Finally, scoring checks of all transformed values were performed.

Both the validation, computation, and interpretation of the QoL- results were done in collaboration with PROMiNET to ensure that the methodological aspects of this data were of high quality. PROMiNET is a Regional clinical research infrastructure for PROMs with the main goals to facilitate increased use of patient-related outcome measures in clinical studies and improving the quality of such assessments (105, 106).

CPRS

The original CPRS was developed by the Swedish psychiatrist Marie Åsberg during the 1970's in collaboration with among others Stuart Montgomery (107). It was a comprehensive survey with 67 items, both self-reported and investigator-scored items. It was estimated to take about 50 minutes to complete and was a mean of scoring a range of psychiatric symptoms both in severity and for change over time. Because of the comprehensive nature, CPRS is not in clinical routine use today. However, since major efforts on validation and interpretation was done (CPRS was translated to both English, German, Italian and Spanish), the usage of items from the original comprehensive version was available. Ten of the questions from CPRS are still of clinical value today in the Montgomery and Åsberg Depression Rating Scale (MADRS) (108). Similarly, Swedish researchers picked out the 17 questions that were assumed to reflect symptoms of hypercalcemia (109). This modified version of CPRS was used in the SIPH-study and other studies on PHPT during the late nineties and through the 2000's. Later, a specific PHPT questionnaire has been developed. Interestingly, the recent disease-specific questionnaire, the Primary Hyperparathyroidism Quality of Life - PHP-QoL, has several similarities with the modified version of CPRS used in the late nineties. As illustrated in Table 2, the thematic similarities are evident and 16 of the 18 items in PHP-QoL are similar to the ones used in the SIPH-study.

Table 2: Comparison of the thematic content of the Modified version of CPRS and PHPQoL:

Questions	CPRS	PHP-QoL
Q1	Stamina/ tiredness	Sleepy
Q2	Mood/ Depression	Felt Weak
Q3	Agitation	Walking difficulties
Q4	Irritation	Shortness of breath
Q5	Sleep	Back pain
Q6	Appetite	Bone/ joint ache
Q7	Concentration	Daily activities
Q8	Decision-making	Leisure activities
Q9	Initiative	Household chores
Q10	Engagement	Irritable
Q11	Pessimism	Depressed
Q12	Worry	Sleep quality
Q13	Unrest	Waking up during night
Q14	Phobias	Consentration
Q15	Obsession	Worry
Q16	Somatic issues	Consentration at work
Q17	Muscle tension/ache	-

At baseline, 176 validated CPRS questionnaires were available for analyses. After 10 years, 120 of the 129 remaining patients completed the CPRS form. Throughout the 10 years, this led to 11,347 individual patient answers with 179 (1.6%) missing values. In the modified version, patients scored each question between zero (best possible quality of life) and seven (worst possible quality of life). CPRS is constructed in a way where each individual answer can give rise to a summary score. Unlike the dimensions in SF-36, it is not possible to calculate a sum score without a complete form, and therefore the number of missing sum scores was higher (70 / 678= 10.3%) than the answers to individual questions. No imputations were done on the missing values in the modified CPRS.

Mortality

In contrast to all the other data in the SIPH-study gathered at the annual visits, mortality data were registry-based data from the Swedish and Norwegian Death Registries. To use registry data for this purpose was already predefined in the original protocol of 1998. An English translated version of the original protocol can be found as a supplement to the paper published in Annals of Internal medicine (110). Since it took 6.7 years to include patients and more than a year to gain access to the data

following the last visit there was a possibility to extend the observation time until July 28, 2018 in Norway and December 31, 2018 in Sweden. This led to an increased mean follow-up time of 15.3 years (SD± 3.2). In this study, we have chosen to include both the 10-year data (primary endpoint) and all available data until death or censoring in 2018. When calculating time to death, all mortality data were specified as a precise date and there were no missing or uncertain timeframes in regards to mortality.

We worked for more than two years with the same applications in Denmark, without getting access to the data, and finally in June 2021, we decided not to do any further attempts to get the mortality data from the Danish Death Registry. Based on their presence at the ten-year visit, all but one of the Danish patients were alive at 10 years. The nine living Danish patients were therefore included in the analyses up until the date of the end of study visit. For the single patient who died, the exact death date and cause of death were available from the clinical CRF and could therefore be included in the analysis in the same manner as the other 43 deaths in the study. This led to a total observation time of 2848 years in the extended observation period.

During the course of the study, PubMed searches on mortality in mild primary hyperparathyroidism have been performed repeatedly. In the work with our paper on mortality, we did a literature search of Clinical trials in English published from inception until September 16, 2021. We searched in PubMed using this search strategy: ("Hyperparathyroidism, Primary"[Mesh] OR "primary hyperparathyroidism"[title] OR primary HPT[title]) AND ("Parathyroidectomy"[Mesh] OR "Parathyroid Glands/surgery"[Mesh] OR "Hyperparathyroidism, Primary/surgery"[Mesh] OR parathyroidectomy[title/abstract] OR parathyroidectomies[title/abstract] OR surg*[title] OR excision*[title]) AND ("Clinical Trial" [Publication Type] OR trial[title] OR study[title]) AND English[lang]. This search string yielded 262 articles on the topic as of September 2021. These articles constituted the basis for the background when working on our publication on mortality in mild primary hyperparathyroidism.

Morbidity

For all patients, a medical history was taken at the yearly visits. Cardiovascular disease, cerebrovascular disease, malignancies and peripheral fractures were specifically investigated at baseline, 2 years, 5 years and 10 years, see Table 1. Data analyses of the morbidity events were based on all existing data throughout the study. The number of patients lost to follow-up was low (n=1 in PTX and n=3 in OBS group). We made an effort during 2018 and 2019 to reduce the number of patients lost to follow-up, and managed to reduce the number to the current four. We also contacted

all study centers directly and gathered missing or unclear data from available records at the individual centers. However, patients who left the study for any of the reasons specified in Figure 6, contributed with their data registered before study discontinuation. As described in the CONSORT flow diagram, Figure 6, there were patients discontinuing the study for various reasons (death, morbidity, etc.). In the case of no event, patients were censored either from the date of study completion or from the date of study discontinuation. Since completing a visit meant that the patient answered the questions assessing medication/ morbidity and fractures in the structured CRF, we had no missing values other than the time censoring stated above.

The start date for all time to event analyses is the same date as the randomization date. In the PTX group, there were no primary or secondary events occurring in the time period between randomization and operation.

Bone Mineral Density

Dual-energy X-ray absorptiometry (DXA) is a rapid and broadly available method to assess Bone Mineral Density (BMD). There are several advantages to DXA for BMD evaluation including; low radiation dose, short scan time, rapid and easy patient setup, good precision, and availability of reliable reference ranges.

In this study, we present BMD measured with DXA at baseline, 5 years, and 10 years. The DXA technology was developed in the late eighties and quickly replaced the current Single photon absorptiometry (SPA) and dual photon absorptiometry (DPA) (111). The two energy sources enabled measurements at sites with variable soft tissue thickness and composition, such as the hip and axial skeleton (112). DXA provides a two-dimensional picture with the result expressed as grams of mineral per square centimeter of bone (g/cm^2). However, interpreting the raw units in g/cm^2 from a user perspective turned out to be difficult. This was both due to different references for different anatomical regions and different machine calibrations from the manufacturers (113). The solution was the introduction of the T-score, which compared the patient's BMD with a mean value of a young healthy reference population and expressed the difference as a standard deviation (SD). In 1994 DXA was established as a gold standard for the diagnosis of osteoporosis by the WHO (114). In women, the DXA values were divided into four different diagnostic categories:
Normal = BMD within 1 SD of a young female reference database.
Low bone mass (osteopenia) = BMD below 1 SD but above 2.5 SD of a young female reference database.
Osteoporosis= BMD below 2.5 SD of a young female reference database.
Severe Osteoporosis (established osteoporosis) = BMD below 2.5 SD of a young female reference database and one or more fragility fractures.

Another value obtained from the DXA measurement is the Z-score. This mathematical score describes the number of standard deviations in BMD the individual differs from an expected mean value based on gender and age. Z-scores are preferred in females prior to menopause and males younger than 50 years (115). Given the age cut-off in the SIPH study, we did not use Z-scores but report on both the absolute values in g/cm^2 and T-score values for all patients.

In the SIPH-study, three different manufacturers contributed to the DXA scans. The majority of scans (n=271) were taken using Lunar Scanners (GE Healthcare). Both Norland (Swissray) and Hologic (Hologic Inc) scanners were used at some centers with a total of 27 and 31 scans, respectively. There are several challenges with comparing DXA-scans across different manufacturers and over time as both hardware, software and reference populations change (116). The SIPH-study group went to great lengths to adjust for both software analysis, reference population, and different hardware throughout the nineteen years of DXA scans. This work led to a methodological publication by Lundstam et al. in 2021 (117).

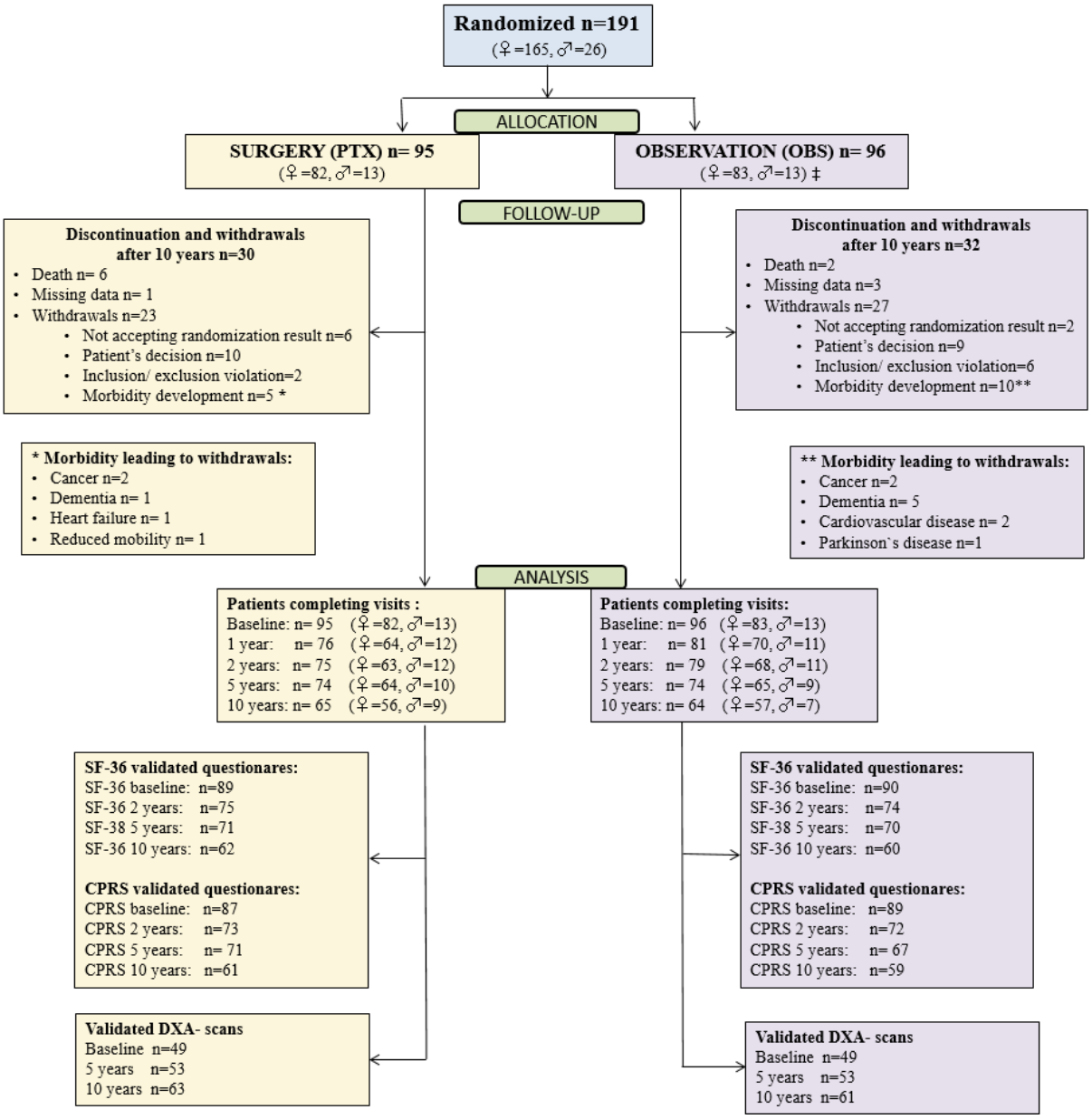

Figure 6: CONSORT (Consolidated Standards of Reporting Trials) flow diagram of patients and available data for analyses presented in this thesis.

Summary of the papers –Results

Paper I: *Effects of Parathyroidectomy on Quality of Life: 10 Years of Data From a Prospective Randomized Controlled Trial on Primary Hyperparathyroidism (the SIPH-Study)*

Observational studies have found a decreased QoL in PHPT with a positive effect of PTX. However, in previous short-term randomized controlled trials, the results have been ambiguous.

In this study, we focused on QoL in patients with mild (previously termed asymptomatic) primary hyperparathyroidism in a prospective long-term randomized controlled trial.

A total of 191 patients were enrolled and randomized for surgical treatment or observation (OBS) in a 1:1 ratio. Two well-validated tools (SF-36 and CPRS) were used at baseline, two, five, and ten years (end of study).

In the SF-36, only the dimension Vitality improved significantly (p=0.017) in the PTX-group compared to the OBS-group. There was a borderline significant difference (p=0.063) in Social functioning favoring the PTX group and a significant improvement within this group (p=0.024).

For the CPRS, several single items improved in both groups over a decade, however, no significant differences were found between the groups. The items that improved in the PTX group were irritation, pessimism, unrest, phobias, somatic troubles, and muscle tension. The items that improved from baseline in the OBS-group were worry, unrest, and somatic troubles. When combined to a CPRS mean sum score there was no difference between the groups.

The observation group had stable or improved values in both SF-36 and CPRS and we concluded that it seems safe to observe mild PHPT for a decade when evaluating only QoL. The improvements in the PTX group were subtle and below what is generally accepted to be clinically significant.

Paper II: *Mortality and Morbidity in Mild Primary Hyperparathyroidism: Results From a 10-Year Prospective Randomized Controlled Trial of Parathyroidectomy Versus Observation*

Several previous papers have assosiated PHPT with increased mortality as well as cardiovascular disease, fractures, and kidney disease. It is still largely unknown if mild PHPT increases the risk of death or morbidity. In this study, we followed 191 patients of whom 96 were randomized to observation and the other 95 were randomized to PTX. They were followed clinically for 10 years with an added registry based follow-up time for mortality (primary endpoint) up until July 2018 in Norway and December 2018 in Sweden.

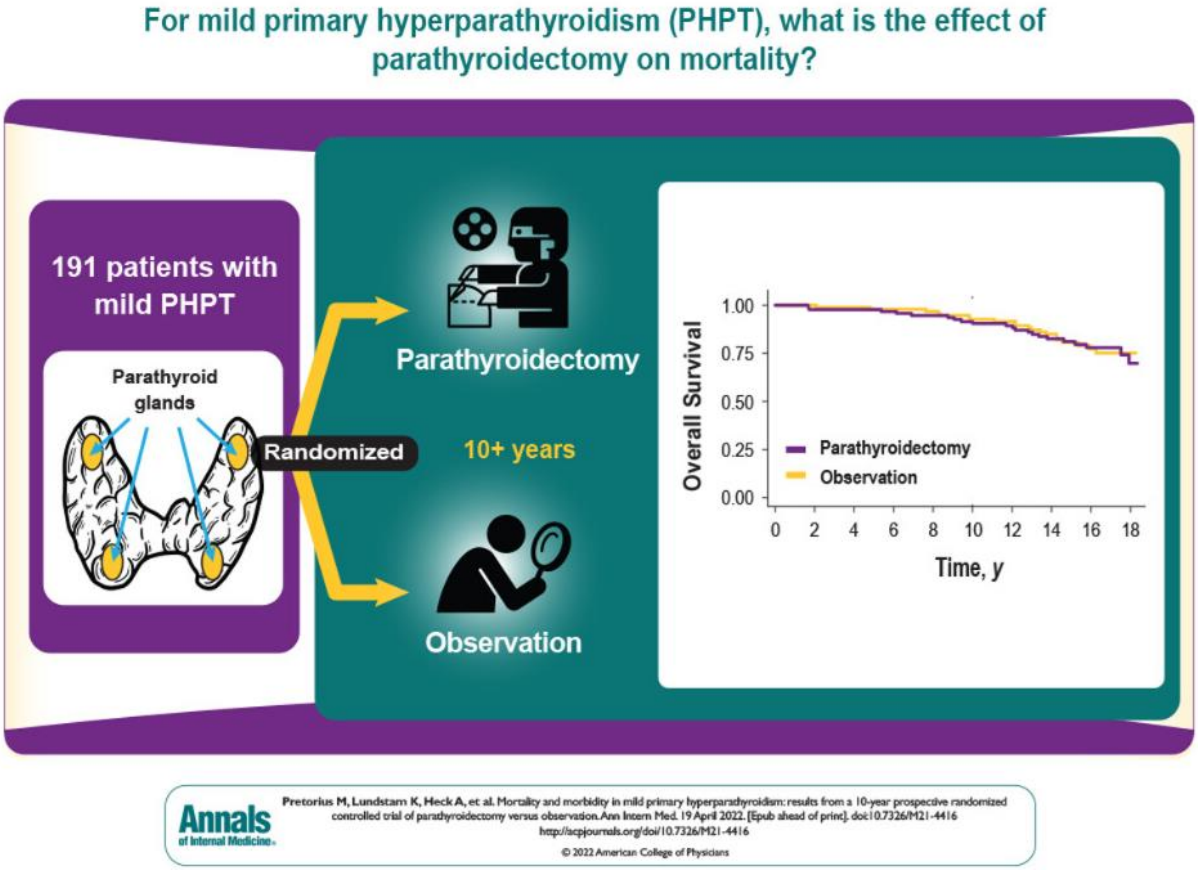

Figure 7: Graphical abstract of mortality, the primary outcome in the SIPH-study. Published in Annals of Internal Medicine (110). Printed with Permission

Fifteen patients died during the 10 years of follow-up. Deaths were evenly distributed in both groups (PTX= 8; OBS= 7). During the added follow-up time 44 patients died with no significant differences between the two groups. The mortality results are also illustrated in the visual abstract (Figure 7).

We report the pre-defined secondary end-points, fractures, cardiovascular events, cerebrovascular events, and cancer. All these were reported annually at study visits. Forty-four new peripheral fractures occurred (PTX = 22; OBS =22), twenty cardiovascular events (PTX= 10; OBS=10), twenty new cancer developments (PTX=13; OBS=7) and seven new kidney stones (PTX=2; OBS=5). Both timing and distribution of the morbidity developments were evenly distributed with no statistical differences between the groups. In addition to the annually registered peripheral fractures, we also report on new vertebral fractures assessed by vertebral fracture assessment (VFA) obtained at baseline, two, five, and 10 years. Fourteen patients developed new vertebral fractures, seven in each group.

The study concludes that PTX does not appear to reduce mortality or morbidity in mild PHPT.

Paper III: *Positive effect of parathyroidectomy compared to observation on BMD in a randomized controlled trial of mild primary hyperparathyroidism*

PTH increases bone turnover and several observational studies have demonstrated a decreased BMD in patients with PHPT. In this paper, we present data on BMD evaluating the effect of PTX within a long-term randomized controlled trial.

In this study, we used mixed model statistics including all validated DXA scans from the 10-year cohort (n=129) from baseline (n=98), 5 years (n=106), and 10-years (n=124).

There was a significant treatment effect of PTX compared to OBS in all DXA-compartments. Lumbar spine (p=<0.001) and femoral neck (p=<0.001) had the largest treatment effect. The OBS-group had a significant decrease in all compartments compared to baseline. The PTX-group, however, had a significant increase in lumbar spine, with stable values in the femoral neck and ultradistal radius and a decline in radius 33%. However, this difference in BMD did not transfer to an increased fracture risk over 10 years. When reviewing the 10-year cohort of 129 patients concerning fractures, 21 of 64 in the OBS-group had an incidental fracture. In the PTX group, 18 of 64 had a fracture with no difference between the groups (p=0.565). We also present biochemical data for the patients who completed the study. Interestingly, the OBS-group had stable calcium levels for a decade without any treatment when excluding the 17 patients who underwent PTX in this group.

Although BMD effects may have a consequence for the patients in an even longer perspective, fractures are of most relevance for the patients. Thus, observation of patients with mild PHPT seems safe also in terms of bone health in a 10-year perspective.

Discussion

Methodological considerations

It is well known that not all research designs can answer research questions on the effectiveness of interventions. This is often described in the hierarchy of evidence, where RCTs are seen as more powerful than cohort series, case-control studies and case series in this regard (118). They are only surpassed in the hierarchy by systematic reviews and meta-analyses of RCTs. Non-randomized and observational trials have inherent biases and limitations, making results on causation unreliable (119). The RCT method was chosen for the SIPH study to assess the effectiveness of PTX. In the case of mild PHPT, this is especially important because there may be considerable selection bias toward surgery. There are both regional and individual differences in regard to which patients are considered to be candidates for operation (120). The final decision on the matter is often done in collaboration with the patient in a form of shared decision-making. This leads to a potential bias where the frail, unwilling patients, or patients with contraindications to surgery, might be excluded from studies.

A key feature of this study is the randomized design. There are several techniques developed for this, including stratified randomization, block randomization, and covariate adaptive techniques. These are meant to ensure the balance of sample size and control for important covariates (121). Even though the hospital center was the only stratification in this study, the groups were well balanced in regard to key factors such as age, gender, morbidities, and calcium levels. Baseline tables are included in all the papers (110, 122, 123). An especially important factor is to ensure concealment in the randomization (124). With simple block randomization, this could be compromised if the investigator is aware of the block size. For example, if a study uses randomization with blocks of four, an investigator who remembers the first three randomized patients will know the result of the fourth patient. This could lead to biased inclusion and is a threat to the randomization method, especially in an open-label study such as the SIPH study. To avoid this, random blocks of between two and ten ensured that the investigator was not aware of the randomization results before randomizing patients.

Power calculations were performed prior to study initiation. Since the natural mortality of mild primary hyperparathyroidism was largely unknown, nine power calculations were done in November of 1994 on the primary endpoint, mortality. The original power calculations were performed at the Department of Clinical Oncology, Rikshospitalet, Oslo University Hospital, Table 3 below. Significance level was set at α=0,05 (5% level), test-strength to 1-β=0,9, 90 %. The clinically meaningful effect of PTX was estimated to 10, 15, and 20 %. Based on evidence at the time, annual

mortality was estimated to be 2.3% with ten-year mortality of 23 % (estimated) (58). Since the estimations were uncertain, there were also calculations for high (30%) and low (20%) mortality. The numbers in each group varied within these estimates between 79 and 389.

Mortality in the group that is observed	Reduction of mortality in the group that is operated	Number of patients in each group
23%	10%, mortality: 13%	305
	15%, mortality: 8%	117
	20%, mortality: 3%	55
20%	10%, mortality: 10%	263
	15%, mortality: 5%	97
30%	10%, mortality: 20%	389
	15%, mortality: 15%	158
	20%, mortality: 10%	79

Table 3: Original power calculations from November 1994.

Based on the power calculations above, the study aimed to include a total of 400 patients and the inclusion period was estimated to be 3 years. However, the recruitment developed slower than anticipated, so at the meeting in the study group on January 25, 2002, only 91 patients so far were included. Based on the low inclusion pace, it was decided to reduce the overall number of included patients to 200. The reasons were:

1) Potential time bias by a prolonged inclusion period, including the – at that time - rapidly increasing use of statins and bisphosphonates
2) Changing international recommendations on management
3) Evolving medical management of complications to PHPT (e.g. clinical introduction of bisphosphonates)

At the following meeting in the study group December 4, 2003, the number of included patients was 142. It was then decided that the aim still was to reach an inclusion of 200 patients, but the end of inclusion was set to the end of June 2005 (6 years and 9 months). Patient no 191 was included in June 2005.

In this study, we chose the Cox proportional Hazard regression model for all time-dependent analyses (mortality and morbidity). Other tests such as the Log Rank test or parametric models could have been chosen. The log-rank test has a very low detection capacity for indicating that one group has a better survival in certain time intervals (125). It does also not provide an estimate of the difference between the groups such as hazard ratio (HR) with confidence intervals, which many journals prefer (126). In the

Cox proportional hazard regression model, there is an assumption that the hazard function only depends on time (127). This assumption can be checked with the Schoenfeld residual-based test for proportional hazard violations (128). In our study, this assumption was met, but if not, several parametric techniques can model survival times (129). However, these methods are often more challenging, and when the hazard functions of the Cox regression models are met this is a safe and proven method and by now the most commonly used for survival analysis (126, 129).

In such longstanding studies, the possibility of recurrent events should be taken into consideration, preferably before the study start and implemented in the study protocol. In the late nineties, the statistical methods for handling recurrent events were not in common use and this was not part of the statistical plan for our study. There are several methods such as the Proportional Intensity Model by Andersen and Gill, the Prentice, Williams and Peterson (PWP) Total time model, or the PWP gap time model. In our material, there were no instances of recurrent events regarding kidney stones. However, one patient had a diagnosis of breast cancer in 2004 with relapse in 2013, both within the study period. We also had a patient with multiple cardiovascular events (heart attack in 2004, angina pectoris in 2006, and heart failure in 2010), all within the study period. Since the number of these recurrent events was low and both could be regarded as a pathophysiological continuum, we chose not to use statistics for recurrent events, and all time-dependent end-points are given as time to first event (130).

As stated in the Table 1 of the original protocol some of the endpoints scheduled in the study have not yet been reported on. This could be exemplified by urinary samples taken at inclusion, two years, five years, and end of study. Unfortunately, standard operational procedures for collecting, sending or reporting of urine samples were not well defined in this study. This resulted in missing samples as well as insecurities regarding this material. We have therefore chosen not to publish data on urinary findings. The presence of a designated study monitor during the entire study could possibly have rectified this in the early phase, but since many of these data and materials were sent to Oslo sometimes with a substantial delay, it was not possible to correct such errors of collection/ handling and analyzing.

There were no reports during the study on “screening failures” or how many patients were contacted during the 6.7 years of inclusion. We, therefore, do not know if those willing to be randomized to the study were different from the not included patients. This could have led to a selection bias, which could limit the conclusions (external validity).

In the later years, more and more evidence has evolved on the pitfalls of using albumin corrected calcium as a diagnostic and prognostic tool. Albumin-adjusted serum

calcium tends to overestimate hypercalcemia and underestimate hypocalcemia when compared to the unadjusted total calcium (131, 132). The probably best measurement is free calcium, which is less affected by binding protein but the use is limited by the fact that it is not readily available outside of the hospital setting and must be analyzed swiftly to avoid pre-analytical errors (133). Free calcium is recommended, when following a patient in the ICU and when the patient has chronic kidney disease (134). Still, when free calcium is not available, recent evidence suggests that total calcium is actually more accurate than albumin-corrected values. When free calcium is available, this would probably be the most accurate test for both diagnostics and treatment decisions (133).

As commented in a recent editorial based on the SIPH study, the nature of conducting a long-term RCT on PHPT with patients randomized to PTX or OBS is challenging and it seems unlikely that a larger or longer RCT will be performed (135).

Ethical considerations

The study was performed in accordance with the Helsinki declaration II and was approved by the regional ethical committees (REC) in all three countries. All participants gave informed consent in their own language prior to inclusion in the study. The consent form included information about the disease, the reasoning behind the study, and what would be expected throughout the ten years of follow-up. It was also clearly stated that patients could at any time leave the study or have an operation (if randomized to OBS-group). Both the risk of progression of disease (skeletal disease, kidney disease, and cardiovascular disease) without operation and the risk of surgical treatment (recurrent laryngeal nerve injury and hypoparathyroidism) were specified.
After two years, an interim analysis was performed on multiple end-points to ensure safety for both groups.

Given the longevity of the study, several additions to the original REC-approval have been sent, both when adding study members and when extending the observation period along with the inclusion of registry data.
In the process of re-applying to the ethical committees in 2018, it was argued that a new patient consent form should be obtained by the remaining patients in the study when applying for the Death Registry data. The nature of this problem is that in order to be included in the registry, consent would have to be obtained from next of kin. Since this application was over twenty years after the first included patients, we argued that this would be inappropriate for the surviving family members. It was also obvious from the original patient consent form and the original protocol that both morbidity

and mortality would be investigated and the approval from REC was given. The REC's in Sweden and Norway approved the extended searches (post-trial) in the mortality registries. There were several reasons for the lack of approval for the ten Danish patients, most of them being of non-ethical nature (mainly issues with the new GDPR rules implemented in the EU in 2018). Still, one reason given in the early phase was that the small number of patients in Denmark would make anonymous extraction of data from the Danish Death registries hard. After arguing that the disease has a high prevalence in this age group, other factors than ethical considerations were the main reasons for not being able to include these registry data.

The study protocol was based on the recommendations from the first international workshop on the management of mild PHPT in 1991 (42). Through the years, these recommendations have been revised; the subsequent workshops (41, 43-45) changed the recommendations concerning diagnostic and therapeutic recommendations. It could be argued that some of the patients in this study would be advised PTX (exemplified by a low BMD) as per these later recommendations. Still, the study clearly demonstrates that hard endpoints are equal after 10 years and retrospectively it would seem that the patients included in the study have not been harmed because of their longtime commitment to the SIPH study.

Based on the original power calculations where the aim was to include 400 patients, it could also be argued that it was unethical to proceed with a long-term study of this magnitude when the original patient number could not be reached. Given the lack of other RCT's and the large population followed for ten years, the data are still relevant today and although this point is applicable, it seems that completing the trial was ethically sound.

Strengths and limitations

The strengths of this study are the randomized, controlled design and the long-term follow-up. Several other observational and registry-based trials include all types of PHPT, while our population is well defined as having mild PHPT. Even though the recommendations for surgery have changed since the first NIH-initiated international workshop in 1991, the majority of the patients in this study would still be classified as mild by today's standards (41).

A major strength of our study is that the original protocol has been meticulously followed throughout the study period, for almost twenty years. As the original protocol was written in the local language, we have translated an English full version and published this with our paper on the primary endpoint mortality (110).

Even though the methods for this study were chosen in the late nineties they were well-validated tools at the time. Most of them are still relevant today such as SF-36, DXA, and laboratory techniques.

There are also limitations. As discussed above, the target number of included patients could not be accomplished. With a 23 % mortality (the mortality after the extended observation period), we cannot exclude reduced mortality to between 15-20% in the PTX group compared to the OBS-group. Moreover, it was not possible to study the mortality in the Danish population beyond the first ten years due to the lack of access to the Danish Death registry data. Nevertheless, the stipulated ten-year mortality could be analyzed.

No power calculations were done for secondary end-points and we cannot exclude a type 2 error based on the secondary end-points. There were also 23 withdrawals in the PTX-group and 27 in the OBS-group. Although these patients were still registered in terms of mortality, they did not contribute with morbidity data after their withdrawal from the study.

The unblinded study design is a potential weakness especially for the QoL-measurements.

Discussion of the results of the separate papers

Paper I: Quality of Life

There is a lack of evidence from long-term randomized controlled trials on the effect of parathyroidectomy on QoL. Although numerous publications have tried to address this in observational or registry-based studies these are only hypothesis generating and cannot answer the question of treatment effect.
The QoL-study showed small improvements in vitality in the PTX group compared to OBS-group and several improvements within the groups in the psychopathological rating scale. Overall, we concluded that the improvements in QoL were subtle with uncertain clinical significance.
There are several strengths and limitations to this paper. First, the study was unblinded for both investigator and patients. This would probably not affect the results of the primary aim (mortality) but could potentially have an important effect on the QoL data. While sham operations (to avoid the placebo effect of surgery) are generally technically possible they should be justified by the clinical question and methodological necessity, since a surgical placebo comes with a risk for the study participants (136). In the SIPH-study it was not seen as feasible to do sham operations.

Sham operations on the neck are not commonly used as it leaves a sizeable, visible scar and two of the most common side effect (wound infection and bleeding) can have a larger risk in the neck area than in other areas of the body. It would also be hard to sustain the blinding over a 10-year period as it would quickly be obvious to the researcher and patient if a patient had an operation, based on the calcium values. The potential placebo effect of operation should be considered, as illustrated by the CPRS data for the initial phase of the study. As illustrated, the long-term follow-up and long time intervals might reduce the placebo effect over time. It is possible that previous short-term studies on QoL in PHPT have underestimated the placebo effect of surgery. Here, we included one-year results and as illustrated in Figure 3 c in the paper, the largest difference between the sum score in CPRS can be found at one year. Hereafter, CPRS scores aligned with time. After 10 years, the scores are almost identical between the groups.
A discussion point is the inclusion of the one-year data. As stated in **Table 1**, QoL measurements was not scheduled for the one-year visit. However, several centers decided to perform SF-36 and CPRS at this time point. This led to 97 SF-36 and 99 CPRS forms being available for the one-year-visit. When using mixed model analysis, all available data could be incorporated into the model, thereby increasing the statistical power. We considered that the data in themselves were interesting, even though not a complete data set from all patients were available. This may have resulted in a bias towards patients willing to answer the questionnaires at this time point. However, we assume that the decision to perform the questionnaire at the one-year-visit was related to the investigator's initiative and not the patients' choice. However, the randomized study design reduced the effect of a potential bias further. We also analyzed all end-points without these one-year data. By that, the results did not change significantly, probably since there were few questionnaires and with a time-dependent model, this one year would only amount to 1/10 of the change from baseline to ten years in the derived mixed model.

The effect on QoL measures was less pronounced in the present study than in comparable non-randomized studies. A potential bias could be a selection bias in observational studies towards patients with the most symptoms at the time, being referred to surgery. This could also be regardless of non-specific symptoms not related to PHPT. Thereby, these symptoms could be relieved by surgery (regression towards the mean).

The use of a generic form such as SF-36 is also debatable. As stated earlier, disease-specific QoL questionnaires for PHPT has been developed after study start of the SIPH-study. Mild PHPT is not known as a disease with a large decline in physical function. This could lead to a ceiling effect, which we saw in both physical dimensions and some of the mental dimensions. Several of the patients (in both groups) scored

maximum points (100) on many of the questions throughout the study. This could be exemplified by question 3i (walking 100 meters) where 87.2% of the patients had top score throughout the study. This could limit the value of the question; however, all the different questions and levels giving rise to a dimension in the SF-36 seem to limit the ceiling effect in our study. Only social functioning had a mean of above 80 at baseline (mean of 81.7) the others ranging from 58.7 (Vitality) to 75.9 (Physical functioning). Thus, the ceiling effect does not substantially limit the values of SF-36 in the SIPH study. The patients in the study were generally quite healthy and well-functioning and it does not seem that the floor effect was present (a large amount of the patients scoring zero thereby limiting the variance so that the independent variable cannot be measured).

In regards to the primary endpoint of mortality and the other four morbidity developments (cardiovascular disease, malignant disease, skeletal disease, and kidney disease) one would rarely argue with multiple testing issues. However, in the QoL paper, there are several end-points. Within the SF-36 there are eight dimensions and seventeen different items in the CPRS along with one sum-score analysis. This sums up to 26 different hypothesis tests and one could easily argue that we should correct for multiple testing. In the case of SF-36, it is generally accepted that these eight dimensions when predefined (not as part of an exploratory or post-hoc analysis) are necessary to evaluate the quality of life without multiple testing corrections. However, the 18 outcomes in CPRS could be considered for multiple testing corrections with for example Bonferroni adjustment. Some argue that corrections for multiple testing is not needed when the comparisons are complementary or when the questions are part of a comprehensive questionnaire (137). In our study, there were two main reasons for not correcting for multiple testing after debating this with PROMINET. Multiple testing corrections are generally used to avoid a type 1 error (false positive result). Our main conclusion from both SF-36 and CPRS was that we could not find clinically significant differences between the groups. Moreover, when using our data for future meta-analyses, it is generally preferable to report on the original data in the study.

The statistical significance does not necessarily mean a clinically meaningful improvement. This was first described as the Minimal Clinical Important Difference (MCID), being “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management” (138). Since, several methods have evolved in describing MCID (139). Generally, the methods are divided into Anchor - or Distribution based methods. Some examples of the anchor-based are: average change, receiver operating curve (ROC)-change, and equipercentile linking. Examples of the distribution-based are standard deviation based (0.5 SD has been suggested), standard error of measurement (SEM), and effect size

(139). Generally, it is better with a disease-specific questionnaire and MCID defined by the relevant population. To our knowledge, no such definition of MCID exists in PHPT. When no predefined MCID exist, a change of 5% to 10% (or in general, 0.5 SD) of the scale breadth is perceptible to patients as a meaningful change (140). Although some studies suggest a threshold of only 5 %, in clinical experience it appears that 10 % is a more reasonable cut-off point, less likely to produce false positives (140).

There are some inherent limitations with self-administered questionnaires. They are limited to literate and native speaking patients, without serious visual problems. They are also not readily achievable with cognitive impairment. Although illiteracy is rare in the Scandinavian countries, cognitive impairment in this kind of long-term study of older patients could be an issue. Generally, this seems to lead to missing values or dropouts of the study. Since we had no continuous measurement of cognitive function, we cannot exclude that this can have had an impact on the answers.

We made efforts to find comparable data from the normal population in all three Scandinavian countries. Even though some normative data exist, it was challenging to find comparable data for both gender-specific, age-specific and from the relevant time periods of inclusion (141, 142). Since the focus of the study was comparison between the two randomized groups, we chose not to include a historical control group. One drawback of this decision was that no SF-36 summary scores could be made. Today it is common to analyze SF-36 data in two main categories; Physical Component Summary (PCS) and Mental Component Summary (MCS). Both are norm-based scoring systems with a mean of 50 and a standard deviation of 10 for the bases of analyses. The approach has the advantages exemplified by the fact that the “ceiling” and “floor” effect (a substantial amount of the patients have a top score of 100 or a score of 0) could be avoided (143, 144). Therefore, we cannot exclude that PCS or MCS, which are common end-points today, could have shown a difference between the groups. Given the data for PCS this seems however unlikely, but since several of the dimensions in the MCS showed improvement in the PTX group a difference cannot be ruled out. There are still debates on the accuracy of the summary scores which makes reporting on the eight classical SF-36 dimensions a validated and safe method (143). A general limitation is also the use of quantitative data (e.g. score 0-100) for data that are qualitative in nature, such as degree of well-being.

Another limitation to be mentioned is the use of the older version of the SF-36. As evident by Figure 5 (content SF-36) there are some questions with quite few items (e.g. bodily pain and social functioning with only two questions). This is not necessarily a problem if the number of levels is adequate. However, some of the critique against version 1 was the fact that some of the questions were dichotomous and thereby giving rise to very few numbers of levels. The most noteworthy is Role-

emotional which only gives rise to four different alternative levels. This was changed in version two with a 5-level response choice making both the role dimensions more robust (145). Changes in these two dimensions especially should be interpreted with caution (146). In the SIPH study, none of these two dimensions were significantly changed but a potential difference might have been demonstrated with more nuance in the questioning. Changes in this dimension will therefore be more exposed to uncertainties, which is why this has been changed in newer versions.

No patients withdrew during the study because of declining QoL, but 19 stated personal reasons (patient's decision) and it cannot be excluded that some of these patients had some sort of symptoms related to psychometric symptoms. However, the number was low and equal in both groups (PTX:10 ; OBS=9). It should be emphasized that the patients were not obliged to give the reason for withdrawal.

An even more recent systematic review on this specific point included seven observational single-arm studies and four RCT's (77). Even though almost all studies had some statistically significant differences, favoring PTX, this study concludes that clinical significance was rarely discussed and five of the studies did not mention the possibility of a placebo effect of surgery. The two-year data from the SIPH trial was one of the included trials. All of the studies were open-label and most of them had a short follow-up.

Paper II: Mortality and morbidity

A discussion point in this study is the use of registry-based data for mortality. The number of patients who died during the study was low, with eight deaths during the ten years of follow-up (PTX=6 and OBS=2). During the conceptualization of the protocol, the inherent problem of keeping elderly patients in the study although they are getting progressively more sick and frail, was anticipated. Therefore, it was decided to use a registry-based end-point, which also meant that the reason for death was made available through registry data. After 10 years, the number of mortalities by these databases almost doubled (n=15). However, the inclusion of the extended mortality period was not protocol based. When extracting the data from the registries we were forced to choose a single date to extract all the data. Since this date had to include all the patients, we chose updated data (the latest time point after the application was approved). Thereby, the dates for the two countries were not equal and it might consequently lead to a different observation time for the two countries. However, the later cut-off date in Sweden for the surviving patients did not give rise to any difference in follow-up time between the two (mean of 16.63 years in Norway and

16.50 years in Sweden). The mortalities in the registry data did not skew the age for the survivors, with the mean age for the survivors at the end of the extended follow-up of 79.80 in Norway and 79.86 in Sweden. When comparing this to the 43 deaths from the Death registries the mean age at death of this group was 83.0 years. This difference probably eludes to the exponential relationship between mortality and age more than randomization in the SIPH-study (147).

The lack of clinical data for the extended observation period is a limitation. Apart from the cause of death, we have no other clinical data for the patients in the Death registries or still alive at the date of data extraction. It is therefore possible that some of these patients from the OBS-group could have had PTX within the extended observation time. However, while this would be clinically relevant, it would not change the statistical result, given all data being analyzed as intention to treat.

Another important point when using data from registries is the accuracy and coverage. The Swedish and Norwegian Death Registries are among the best-performing registries worldwide (148). The Norwegian Causes of Death Registry is described in this way in the publication; Data quality in the Causes of Death Registry (149), "*Degree of coverage refers to the proportion of the population encompassed by a registry. Completeness indicates whether information is actually collected from the persons who are included in the registry. In Norway, both of these are high; on the whole, we assume that we have medical information on more than 98 % of all deaths. The Causes of Death Registry encompasses all residents, irrespective of whether they die in Norway or abroad, and since 2012 non-residents (tourists, labour migrants etc.) who die in Norway are also included. Cross-checks against the National Registry are made, and reminders are sent to the Chief Municipal Medical Officers to collect missing certificates*". As described, there are limitations concerning underlying and secondary causes of death. However, we were aware of this limitation and have not included such data in our study. All our data on morbidities were prospectively gathered from the study, and not from registry data. The Swedish National Board of Health and Welfare is responsible for the Swedish Cause of Death registry. According to their report; Dödsorsaksstatistik. Historik, produktionsmetoder och tillförlitlighet (https://www.socialstyrelsen.se), (English translation of the title would be "Cause of death statistics – history, production methods, and reliability"), all deaths reported to the Swedish Tax Agency are included in the Cause of Death registry. This means that no deaths are missed, even though the cause of death can be missing for some small numbers. However, in our study, there were no missing causes of death for any of the Swedish or Norwegian patients. Consequently, the degree of both coverage and completeness is high (above 98 %) in both countries (148, 149).

As mentioned in the introduction, several studies have shown increased mortality in PHPT. Of special interest is a recent Swedish publication (150) showing an increased risk of death, fractures, and cardiovascular disease. Interestingly, the majority of the patients in this study were recruited and followed in the same country (Sweden) and same period (2006-2017 vs 1998-2017), as the SIPH study. Thus, time or geographical factors should not be a major bias here. There are numerous inherent limitations to a registry-based, retrospective cohort study. Specifically, it is difficult to identify differences between a hospital-derived group of PHPT patients and the age, sex, and residency-matched presumably healthy control group. Both groups in this recent study were scored by the Charlson Comorbidity Index (151). Although commonly used to adjust for covariates, it is not without limitations (152). In order to compare the groups, the coding practice is assumed to be equal in a hospital and the broader primary community. One example of where the comorbidity scale does not fully encompass differences, is with medication. In the study, the hospital group was using more of all medication groups during the last year (except for calcium supplementation and drugs against dementia). This is also the case for medication that presumably would not be used to treat either PHPT or the complications. This could be exemplified by a 10.0 % Prednisolone use in the PHPT group and 6.9 % in the Control group, moreover a 10.8 % opioid use in the PHPT group versus 6.7% in the Control group. The potential biases comparing two such groups are many and hard to quantify through retrospective registry data. When comparing the PTX-group to conservative treatment it must be commented that the groups were different before adjustment for covariates. In this study, the conservative group was almost ten years older (71.1 vs. 62.5 years), and this would presumably affect many other factors not easily corrected for by just age, sex, and the Charlson Comorbidity Index. The fact that all hazard ratios dropped after adjustment for covariates supports that the relationship could be due to biases. Lastly, since the ICD-10 diagnosis E21.0 does not include any form of severity score, it is not known how many of the included patients in registry studies met the surgical criteria (exemplified by calcium >0.25 mmol/l or symptoms of hypercalcemia). Current guidelines recommended PTX for patients with osteoporosis, kidney disease, symptoms of hypercalcemia, or calcium values above 0.25 mmol/l. The fact that many studies, including this recent cohort study, do not identify or exclude patients with severe disease or organ manifestations, who meets these criteria for surgery makes interpretation difficult.

Morbidity:

We found no significant differences with regard to development of morbidity during the ten years of study duration. There were 1723 completed visits in the SIPH-study. In all of these visits, incidental disease was examined anamnestically.

No power calculations were done for CVD, malignancy, fractures or kidney stones. It is possible that these secondary endpoints were not powered to exclude a type 2 error (153), especially if few patients experience the end-point during the study. This could be exemplified by symptomatic kidney stones with a total number of seven cases (five in the OBS-group and two in the PTX-group) giving a HR-ratio of 0.34 but with a wide confidence interval of 0.06 to 1.82 (non-significant). Given the large confidence interval on this morbidity event a considerable risk for type 2-error might exist. However, the low frequency of the event over a duration of ten years is still interesting. Today, prior kidney stones is an indication for PTX. None of the five patients who had a kidney stone prior to inclusion in the OBS-group suffered from symptomatic kidney stone during the study. The low incidence of kidney stones over ten years indicates that this is not a common clinical problem in conservatively treated mild PHPT. Further efforts to identify patients with mild PHPT who are at risk of developing kidney stones and renal disease should be made in the future.

Kidney function was one of the predefined secondary endpoints. Serum creatinine levels were one of the biochemical tests taken at every visit (along with calcium, PTH, albumin, and hemoglobin). Creatinine levels were similar in both groups at baseline. Over the ten-year period, the serum creatinine levels decreased significantly in both groups. In the PTX group from 81.4 (± 17.4) µmol/L to 74.9 (± 16.5); ($p= 0.011$) and in the OBS group from 79.9 (± 13.9) µmol/L to 70.9 (± 18.1); ($p < 0.001$). After ten years, however, there was no significant difference between the groups ($p=0.218$). The equal decline in serum creatinine (within the reference range) in both SIPH study groups could be due to a gradual decrease in muscle mass in this age group or a selection bias with the possibility of patients with higher creatinine levels adhering less to the protocol. We did not use estimated GFR, i.e. taking not only serum creatinine, but also gender, age and weight into account, which better reflects the actual renal function (measured GFR) in an individual. In recent years, the use of Cystatin C has been introduced clinically, as an alternative or complement to serum creatinine, as it is independent of the muscle mass. Nevertheless, with the patients being their own controls over time in the SIPH study, it still seems reasonable to use serum creatinine as a marker of renal function. It would be of interest if there were differences between the groups or if any of the groups increased in s-creatinine. Clearly, this was not the case in our data. Since paper II focused on symptomatic disease, we chose not to include these laboratory end-points in this paper.

There is still a concern about cognitive impairment and the risk of dementia related to PHPT. Several case reports exist on PHPT as a reversible cause of dementia (154-156). The SF-36 and CPRS do not specifically address cognitive domains and we had no structured cognitive evaluation during the study. At the end of the study, five patients had left the study because of dementia in the OBS-group versus one in the

PTX-group. Therefore, we chose to search for incidental dementia in the CRF of each patient. Two more patients were diagnosed with dementia during the study, both within the PTX-group. Since this was not a predefined end-point and we did not have any structured way of evaluating cognitive failure, we chose not to include these findings in the paper.

We chose to combine all the clinical morbidity events in the paper leading to a total of 101 predefined events (cardiovascular events, fractures, cancers, and kidney stones). Scientifically, this is controversial when the events do not share a common pathophysiology. The development of kidney stones and cancer are examples of two different diseases. However, we did this to demonstrate that the presumed morbidity events of untreated PHPT were almost identical in the two groups. In the PTX-group there were 52 events and in the OBS-group 49. Even though there were few events in some of these end-points, the total burden of potential complications to PHPT seems to be equal in the two groups. This is further supported by the fact that all the disease-specific causes of death in the groups were almost identical in both numbers and distribution (Table 2, paper II).

As commented in a recent editorial on the SIPH-study the results provide a strong rationale for non-operative management of mild PHPT (157). The guidelines from the international workshops, although high profile, is subject to the expert opinion biases. The lack of clinically relevant effects on clinical outcomes suggests that many patients can be managed without operation. The SIPH study also demonstrates that RCT's on surgery vs. observation are feasible and can provide important clinical information.

Paper III: Bone Health

The effects of PTH on bones are complex and still not completely understood. Especially the different effects on cortical and trabecular bones have been subject to several publications (158, 159). Approximately 80 % of the bone mass is categorized as cortical with the remaining being trabecular compartments. Trabecular bones have a lower calcium content, larger surface exposed to the bone marrow and bone turnover is generally higher than in the cortical bone (160). Previous studies have demonstrated a well preserved and even improved trabecular bone density but increased cortical porosity and thinning in PHPT (159, 161). The difference in the two types of bone could be explained by the complexity of both the anabolic and catabolic effect of PTH on bone. The anabolic effect is through stimulation of mesenchymal stem cells and osteoprogenitor cells to form new osteoblasts, as well as decreasing the production of sclerostin. In addition, PTH increases the release of osteoblastic IGF-1 and FGF2

having proliferative and anti-apoptotic effects on osteoblasts (162). However, the net effect of continuous PTH stimulus is to release more calcium into the circulation and thereby increasing bone resorption more than formation. This catabolic effect seems to be mainly through enhancing production of RANKL and inhibiting its antagonist OPG leading to increased osteoclastic bone resorption (162). Both these processes increase bone turnover in a way that seems to diminish cortical bone but sparing trabecular bone. In our study, the DXA-values or the treatment effect of PTX does not correlate to fracture incidence and it would seem that there is a need for even better understanding of which patients with PHPT have the highest fracture risk.

Even though several studies show a lower T-score in the cortical sites, the reversibility of the cortical bone loss with PTX has not been demonstrated (163, 164). The difference in these two compartments could be explained by the different PTH effects. As seen in treatment with Teriparatide this leads to an increase of especially trabecular bone (165). In contrast to previous studies, our study clearly demonstrates that there is a treatment effect also in the typical cortical bones (Radius 33% has about 95 % cortical bone (166))

A major issue when discussing mild or asymptomatic PHPT is the lack of a clear definition of mild disease. This has led to heterogeneity in the studies on the topic. In terms of bone disease, the original protocol for the SIPH study stated that only hyperparathyroid bone disease (osteoitis fibrosa cystica generalisata) was an exclusion criterion. The classical and pathognomonic bone disease of PHPT, Osteitis Fibrosa Cystica with Brown tumor is today a distinct rarity (167, 168). The protocol specified that typical osteoporotic fractures e.g. in lower arm, hips and spine were not considered as hyperparathyroid bone disease and patients with low BMD on DXA were not excluded from the study. During the twenty-four years between the original protocol and the latest international guideline on management, the definition of mild PHPT has changed several times. The lack of a clear definition has been a focal point of the 2022 Fifth International Workshop on PHPT (41). In the recent paper on classical and non-classical manifestation of PHPT it is proposed to refine the definition of asymptomatic PHPT into two categories (168). The first category is asymptomatic PHPT without target organ involvement. The second is with evidence of target involvement. To make the classification more universal, it was suggested that evaluation of target involvement should include the following: Inquire about fracture history, BMD (spine, hip and forearm) and imaging to rule out vertebral fractures. For asymptomatic kidney disease, e-GFR or creatinine clearance, kidney imaging studies and 24-hour calcium excretion is recommended. This refinement would, if used uniformly, lead to better characteristics and improve the assessment and development of future evidence based guidelines. If these recommendations had been applied on the SIPH-study patients at baseline with regards to the skeletal involvement, the following

number of patients with a refinement of the diagnosis to asymptomatic PHPT with evidence of skeletal organ involvement would be given:

Patients with fracture history: n= 51 (22 in PTX and 29 in OBS)

Patients with T-score < -2.5: n= 34 *

Lumbar Spine below -2.5: n= 24 (10 in PTX and 14 in OBS)

Femoral neck below -2.5: n= 11 (8 in PTX and 3 in OBS)

Radius 33% below -2.5: n= 15 (4 in PTX and 11 in OBS)

Patients with baseline vertebral fracture: n= 9 (5 in PTX and 4 in OBS)

*The number of target organs involved is higher than the number of patients since several of the patients presented more than one skeletal involvement.

Since no detailed anamnestic history was taken after finding a radiological vertebral fracture or peripheral fracture, it is not known how many of these would be classified as low-energy trauma at baseline. Even though several of the patients in the SIPH-study would have been advised PTX by today's standard, it is still interesting to see that most of the hard end-points were not increased in the OBS-group after ten years. This further emphasizes the need to continuously evaluate the current understanding of PHPT being a progressive disease and that PTX had a significant effect on skeletal hard end-points, not supported by our data. The updated proposal of the 2022 International Workgroup is an important step towards standardization and more homogenous patient populations.

A recent systematic review and meta-analysis included 12 studies after identifying 804 papers on the topic (169). An increased risk of fracture was found in PHPT patients (risk of any fractures (OR: 2,01), forearm (OR: 2.36) and spine (OR: 3.00)). Four studies included patients with mild PHPT and showed an increased risk of vertebral fractures of 4.22 with a confidence interval of 2.20 to 8.12. There are limitations to the analysis, as all of the studies were observational (eight cross-sectional and four cohort studies). There was a high heterogeneity in the studies and the methods of diagnosing vertebral fractures were not uniform and could underestimate the true number of vertebral fractures. The huge span of prevalence in vertebral fractures from 1% (170) to 46 % (171) indicates that vertebral fracture assessment in patients with PHPT was done differently across the included studies or differences in patient selection.

In the SIPH study, vertebral fractures were assessed at plain x-rays of the thoraco-lumbar spine. It should be acknowledged that computerized tomography (CT) potentially could reveal vertebral lesions not detected at plain x-rays. Using low-dose CT would have been preferred if the design of the study was done today. In any case,

the same x-ray technique was used for both PTX and OBS groups, making the groups comparable, although at a slightly lower diagnostic level.

Because of the unclear relationship between BMD and the risk of fractures, other modalities have been proposed in order to give a better estimate of fracture risk in PHPT. Trabecular bone score (TBS) can be obtained by DXA measurement and describes the trabecular microarchitecture (172). The possibility for evaluating fracture risk with TBS in PHPT was introduced with the fourth international workshop in 2014, however results since then have been conflicting and TBS is not recommended in the current 2022 guidelines (173, 174) . High resolution peripheral quantitative computed tomography HRpQCT is an imaging modality which can both assess volumetric bone mineral density and describe the microarchitecture of cortical and trabecular bone (175). Studies with HRpQCT have demonstrated a decrease in bone mineral density (both in cortical and trabecular space) as well as reduced cortical thickness and trabecular number in PHPT compared to controls (176, 177). This again confirms that the trabecular bone is affected in both symptomatic and mild PHPT. HRpQCT is not readily available for clinical use, is time consuming and has a higher radiation dose than DXA. A more direct way of measuring bone strength is through impact microindentation (178). This technique has also been investigated in PHPT with significantly lower bone material strength index in patients with primary hyperparathyroidism compared to controls (179). In addition, this study also show significantly lower Bone Material Strength index (BMSi) in PHPT patients with fractures compared with PHPT without fragility fractures. Neither TBS, HRpCT nor impact microindentation were available at the time of planning the SIPH study.

A major limitation with gathering DXA-data over a time span of almost 20 years, is the inevitable updates in hardware, software, reference population and analysis settings. Three manufacturers (Lunar, Norland and Hologic) were used in the SIPH-study. It is well established that DXA-scans from different manufacturers is a limitation (180). It is also known that hardware from the same company can pose a threat when analyzing longitudinal data (116). A majority (82%) of the scans in the SIPH-study were performed with Lunar technology. Even when only analyzing scans from Lunar, 6 different hardwares, 26 different softwares and 16 different reference populations values were used during the trial period leading to an array of potential sources of error. Some of these are well known and described (181, 182). The limitations are exemplified by the implication of the NHANES/Lunar Femur reference population (version 111, 112 and 113) leading to a difference in the relationship between the BMD and T-score. This has been previously described but since it mostly affects the high area of the T-scores it has had lower implications in the area of clinical

relevance (183). This change in the reference population can be clearly seen also in the SIPH-study as illustrated in Figure 2b in the recent paper of Lundstam et al (184).

When validating the DXA-data for this publication, it became evident that discrepancies in the longitudinal set up could have led to substantial misinterpretations and wrongful comparisons with former scans due to change of software. In the forearm compartments two distinct patterns of BMD-T-score relationships emerged. Most of the differences between the two could be explained by a single calibration setting. The setting Single Photon Absorptiometry (SPA)-calibration was changed during the study period to "Lunar"-calibration leading to a 24% increase in BMD values without affecting the T-scores. In collaboration with Lunar, we published these data and were also provided with a conversion formula by the company; BMD (SPA) = BMD (Lunar) x 0.804 (184). With this conversion the BMD-T-cores aligned in the measurements of radius 33% as illustrated in Figure 8a and 8b.

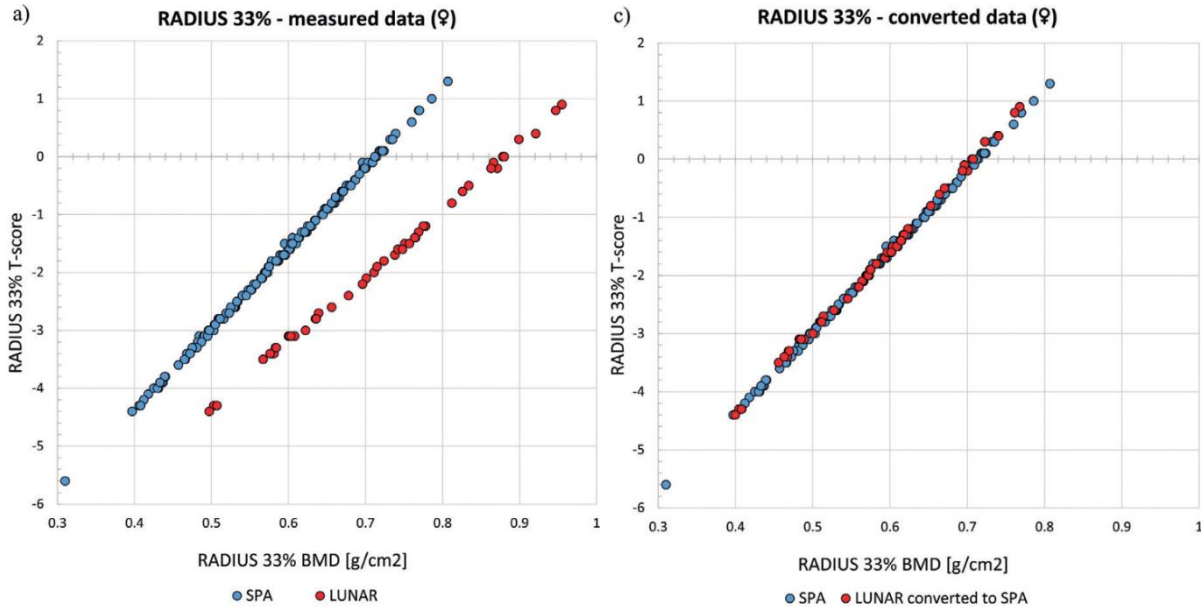

Figure 8: a) Illustration of impact of the two settings; SPA-calibration blue dots and Lunar calibration (red dots) b) Same measurements after conversion of the data. Published in Journal of Clinical Densitometry. Published with permission.

The different impact on cortical and trabecular bone and differences in the bone compartments should theoretically be irrelevant, when measuring total body (TB). Total body should be a better measurement of the patient's cortical bone density as the measured area is large with less sources of error, compared to other compartments. However, this technique has not been without limitations (185). For the TB DXA-data,

we demonstrated a difference when the manufacturer changed from Basic to the Enhanced analysis setting. This gave rise to two different BMD-T-score relationships. Thirty-six scans were available for reanalysis (at Oslo University Hospital) and when reanalyzing the scans with the same software settings much of the difference could be accounted for with the two groups aligning. Unlike the errors with the reference populations, the largest discrepancies were in the lower values, of relevance for clinical decision making. Because a majority of the scans was not available for such reanalysis, and there were still some unresolved issues with the long-term data for TB, we chose not to include the data in our publication on the 10-year data. This decision was also based on a clinical argument. While TB-DXA is the preferred method of measuring body composition, its utility in assessing BMD in patients with PHPT has not been established (186). Both in assessing osteoporosis in general and in PHPT, only lumbar spine, hips and radius has been among the indications of PTX.

Another discussion point with the DXA and fracture data was the allowance of osteoporotic therapies. Bisphosphonates were not in routine use at the time of protocol writing. During the study period, the use of bisphosphonates increased. All medical therapies were recorded in the CRF at baseline, two, five and after ten years. For the 191 patients included in mortality and morbidity data analyses, we found no statistical difference between the two groups for the use of bisphosphonates (6 in PTX vs 15 in OBS; p= 0.08). However, when analyzing only the 129 patients for this paper, 18/129 patients used bisphosphonates at any time point. There was a significant difference between the groups with 5 in the PTX-group and 13 in the OBS-group; p = 0.039. This could have affected both the DXA and fracture data. Given that more patients in the OBS-group received bisphosphonates you would suspect that this would increase the BMD in this group. Since the result showed the opposite with a significant treatment effect in all compartments, it seems unlikely that this could change the conclusion of the DXA-results for the study. The low fracture rate in the OBS-group could to some extent be affected by the use of bisphosphonate. Several studies have demonstrated a decrease in fracture rate with the use of bisphosponates in PHPT. However, in the dataset used for the fracture analyses (including all 191 patients) there was no significant difference in the use of bisphosphonates. Still, if a similar study protocol would have been formed today, the use of bisphosphonates in the OBS-group would be defined and controlled.

The optimal timing and dosage of combined bisphosphonate treatment with parathyroidectomy is not well established. A retrospective cohort study concluded that bisphosphonate initiation after PTX might interfere with the beneficial effects of PTX (85). However, a recent randomized controlled trial demonstrated a significantly higher BMD in the group with surgery combined with postoperative Zoledronic acid

(86). However, case reports suggest that this could increase the risk of postoperative hypocalcemia (87).

Another bone active therapy that was allowed, both at baseline and during the study, was estrogen. Twenty-two of the 129 in the DXA cohort had estrogen therapy at any time point during the study when including both per oral and transdermal estrogens. There was no difference between the two groups in estrogen use (13 in PTX and 9 in OBS; p=0,370). Unlike bisphosphonates, estrogen therapy was frequent at baseline (26 of 32, respectively, in the SIPH-cohort) and only six patients started with estrogens during the study follow-up.

SERMs were also allowed and a total of four patients (two in each group) used this during the study. The low number and equal distribution makes SERM unlikely to influence the results.

Conclusions and implications

Parathyroidectomy remains the treatment of choice in overt or severe PHPT. Our literature search showed that data on the effect of PTX with regards to mortality and morbidity in mild PHPT was sparse and that the results from observational and retrospective studies were conflicting. Our findings suggest that observation without operation is safe in terms of both survival and key morbidities, such as cardiovascular disease, fractures, kidney disease, and malignancies for several years.

The SIPH study is a landmark study for the management on mild PHPT. The study is the largest and longest lasting RCT on the topic. So far, 12 papers have been published on these data. Since the second international workshop on PHPT, interim data from the SIPH-study has been cited frequently in all the international workshops and European consensus statements. The three publications included in this thesis are based on end of study data. Only the QoL paper was published in time to be included for the 5th international workshop in 2022 and was well acknowledged. The decision on not to include QoL-measurements as indication for PTX was probably in part due to the new 10 year data from the SIPH-study. The mortality and morbidity data was published after the conclusion of data gathering for the fifth international workshop. However, these results have impact for future guidelines, and provide evidence for shared decision making with patients.

The DXA-paper has been published in January 2023 (123). The clear decline in BMD-values in the observation group could lead to an even greater focus on patients with fractures or with a high risk of fractures. However, the low and equal fracture incidence over a decade seem to advocate for observation without treatment in patients without organ involvement.

Although tedious, the SIPH study has demonstrated that it is possible to carry out long term RCT's to answer many of the still unanswered clinical questions around mild PHPT. Generally, the indications for PTX in mild PHPT have increased over time. The data from the SIPH-study support the safety of observation without PTX in a ten-year perspective.

Future Perspectives

Although PHPT is a common endocrine disorder, the long-term natural course in mild disease is still largely unknown beyond 10 to 15 years. The changing clinical facets of this disease during the last decades have led to an array of different studies with different severity of PHPT. The suggestion of the last international workshop to further define asymptomatic disease into two separate categories and to standardize target organ involvement makes future studies with well-defined groups possible. Both asymptomatic PHPT without organ involvement and normocalcemic PHPT are groups in need of further studies, preferably in a randomized controlled setting.

The unclear correlation between DXA values and fractures is a concern and studies to predict fracture risk in mild PHPT is needed. The role of the new techniques for evaluation of biomechanical properties of bone, such as TBS, HRpQCT, FRAX and biochemical bone markers of bone turnover needs further elaboration in PHPT.

The timing and efficiency of other bone active treatments in relation to PTX for PHPT is still unknown. Many patients continue to have low bone density after PTX and the timing of bisphosphonates, Denosumab and Romosozumab in these patients will be of interest for further research.

The low frequency of nephrolithiasis and kidney failure in our study highlights the need to develop tools to understand the risk of renal manifestations.

With updated current guidelines in hand, there is also a need to evaluate barriers in compliance and knowledge of international guidelines.

The current calcium-threshold of 1 mg/dl (0.25 mmol/L) above the reference range is arbitrary and further studies to evaluate the optimal calcium value for PTX are needed. The role of free calcium and whether, or how to, use albumin-correction for calcium levels needs clarification.

The single patient included in the SIPH study with parathyroid carcinoma highlights the need for better diagnostic tools to separate this rare, but important disease from its common counterpart of PHPT.

In terms of the neurocognitive effects of PHPT, there is a further need to develop and validate disease specific methods. These tools should capture typical symptoms and signs of the disease in a way that reversibility by PTX could be demonstrated, if present. Even though such a questionnaire has recently been developed, it is still necessary with further development, validation and translation, as well as usage in a proper randomized controlled setting. The PHPQoL does not include cognitive functions and this important aspect was not investigated in the SIPH-trial. Studies to

evaluate cognitive impairment, especially in older individuals with PHPT is lacking. If mild PHPT leads to a cognitive decline over time this would change the management of mild PHPT.

References

1. Potts JT. Parathyroid hormone: past and present. J Endocrinol. 2005;187(3):311-25.
2. Khan M, Jose A, Sharma S. Physiology, Parathyroid Hormone. StatPearls. Treasure Island (FL): StatPearls Publishing

3. Martin TJ, Sims NA, Seeman E. Physiological and Pharmacological Roles of PTH and PTHrP in Bone Using Their Shared Receptor, PTH1R. Endocr Rev. 2021;42(4):383-406.
4. Silva BC, Bilezikian JP. Parathyroid hormone: anabolic and catabolic actions on the skeleton. Curr Opin Pharmacol. 2015;22:41-50.
5. van Abel M, Hoenderop JG, van der Kemp AW, Friedlaender MM, van Leeuwen JP, Bindels RJ. Coordinated control of renal Ca(2+) transport proteins by parathyroid hormone. Kidney international. 2005;68(4):1708-21.
6. Brenza HL, Kimmel-Jehan C, Jehan F, Shinki T, Wakino S, Anazawa H, et al. Parathyroid hormone activation of the 25-hydroxyvitamin D3-1alpha-hydroxylase gene promoter. Proc Natl Acad Sci U S A. 1998;95(4):1387-91.
7. Bacic D, Lehir M, Biber J, Kaissling B, Murer H, Wagner CA. The renal Na+/phosphate cotransporter NaPi-IIa is internalized via the receptor-mediated endocytic route in response to parathyroid hormone. Kidney international. 2006;69(3):495-503.
8. Rendina-Ruedy E, Rosen CJ. Parathyroid hormone (PTH) regulation of metabolic homeostasis: An old dog teaches us new tricks. Mol Metab. 2022;60:101480.
9. Lupp A, Klenk C, Röcken C, Evert M, Mawrin C, Schulz S. Immunohistochemical identification of the PTHR1 parathyroid hormone receptor in normal and neoplastic human tissues. European journal of endocrinology. 2010;162(5):979-86.
10. Usdin TB, Gruber C, Bonner TI. Identification and functional expression of a receptor selectively recognizing parathyroid hormone, the PTH2 receptor. The Journal of biological chemistry. 1995;270(26):15455-8.
11. Baczynski R, Massry SG, Magott M, el-Belbessi S, Kohan R, Brautbar N. Effect of parathyroid hormone on energy metabolism of skeletal muscle. Kidney international. 1985;28(5):722-7.
12. Baczynski R, Massry SG, Kohan R, Magott M, Saglikes Y, Brautbar N. Effect of parathyroid hormone on myocardial energy metabolism in the rat. Kidney international. 1985;27(5):718-25.
13. Bagó AG, Dimitrov E, Saunders R, Seress L, Palkovits M, Usdin TB, et al. Parathyroid hormone 2 receptor and its endogenous ligand tuberoinfundibular peptide of 39 residues are concentrated in endocrine, viscerosensory and auditory brain regions in macaque and human. Neuroscience. 2009;162(1):128-47.
14. Bollerslev J, Pretorius M, Heck A. Parathyroid hormone independent hypercalcemia in adults. Best practice & research Clinical endocrinology & metabolism. 2018;32(5):621-38.
15. Bollerslev J, Schalin-Jäntti C, Rejnmark L, Siggelkow H, Morreau H, Thakker R, et al. MANAGEMENT OF ENDOCRINE DISEASE: Unmet therapeutic, educational and scientific needs in parathyroid disorders. European journal of endocrinology. 2019;181(3):P1-p19.
16. Owen R. On the Anatomy of the Indian Rhinoceros. Trans Zool Soc Lond. 1852;4: 31-58.
17. IV S. Omen ny Kortel hos mennisken och atskilige baggdjur. Lakarefore rings (ed) Upsala,. 1880:pp; 441-71.
18. Gley E. Sur les fonctions du corps thyroide. CR Soc Biol. 1891;43:841.
19. Collip JB. The extraction of a parathyroid hormone which will prevent or control parathyroid tetany and which regulates the level of blood calcium. Journal of Biological Chemistry. 1925;63(2):395-438.
20. Hegele RA, Maltman GM. Insulin's centenary: the birth of an idea. The lancet Diabetes & endocrinology. 2020;8(12):971-7.
21. von Recklinghausen F. Die fibrose oder deformierende Ostitis, die Osteomalacic und die osteoplastische Carcinose, in ihren gegenseitigen Beziehungen. Rudolf Virchow Festschriften. 1891:1-89.

22. Cook M, Molto E, Anderson C. Possible case of hyperparathyroidism in a Roman period skeleton from the Dakhleh Oasis, Egypt, diagnosed using bone histomorphometry. Am J Phys Anthropol. 1988;75(1):23-30.
23. Rowlands BC. Hyperparathyroidism: an early historical survey. Ann R Coll Surg Engl. 1972;51(2):81-90.
24. Kafetzis ID, Diamantopoulos A, Christakis I, Leoutsakos B. The history of the parathyroid glands. Hormones (Athens). 2011;10(1):80-4.
25. Albright F RE, Jr. Baltimore, MD. The parathyroid glands and metabolic bone disease, selected studies. : The Williams and Wilkins Company; 1948.
26. Potts JT, Jr., Murray TM, Peacock M, Niall HD, Tregear GW, Keutmann HT, et al. Parathyroid hormone: sequence, synthesis, immunoassay studies. The American journal of medicine. 1971;50(5):639-49.
27. Jüppner H, Abou-Samra AB, Freeman M, Kong XF, Schipani E, Richards J, et al. A G protein-linked receptor for parathyroid hormone and parathyroid hormone-related peptide. Science. 1991;254(5034):1024-6.
28. Lundgren E, Rastad J, Thrufjell E, Akerström G, Ljunghall S. Population-based screening for primary hyperparathyroidism with serum calcium and parathyroid hormone values in menopausal women. Surgery. 1997;121(3):287-94.
29. Minisola S, Arnold A, Belaya Z, Brandi ML, Clarke BL, Hannan FM, et al. Epidemiology, Pathophysiology, and Genetics of Primary Hyperparathyroidism. Journal of Bone and Mineral Research.n/a(n/a).
30. Yeh MW, Ituarte PH, Zhou HC, Nishimoto S, Liu IL, Harari A, et al. Incidence and prevalence of primary hyperparathyroidism in a racially mixed population. J Clin Endocrinol Metab. 2013;98(3):1122-9.
31. Siilin H, Rastad J, Ljunggren O, Lundgren E. Disturbances of calcium homeostasis consistent with mild primary hyperparathyroidism in premenopausal women and associated morbidity. J Clin Endocrinol Metab. 2008;93(1):47-53.
32. Insogna KL. Primary Hyperparathyroidism. The New England journal of medicine. 2018;379(11):1050-9.
33. Bhadada SK, Arya AK, Mukhopadhyay S, Khadgawat R, Sukumar S, Lodha S, et al. Primary hyperparathyroidism: insights from the Indian PHPT registry. Journal of bone and mineral metabolism. 2018;36(2):238-45.
34. James R, James J, Vij AS, Vij KK. Spontaneous fractures in a 30-year-old woman following a simple fall. BMJ case reports. 2013;2013.
35. Garingarao CJ, Paz-Pacheco E, Jimeno CA. Primary hyperparathyroidism from a probable ectopic parathyroid adenoma with severe skeletal disease and vitamin D deficiency. BMJ case reports. 2014;2014.
36. Bilezikian JP, Bandeira L, Khan A, Cusano NE. Hyperparathyroidism. Lancet (London, England). 2018;391(10116):168-78.
37. Singh DN, Gupta SK, Kumari N, Krishnani N, Chand G, Mishra A, et al. Primary hyperparathyroidism presenting as hypercalcemic crisis: Twenty-year experience. Indian J Endocrinol Metab. 2015;19(1):100-5.
38. Elvius M, Lagrelius A, Nygren A, Alveryd A, Christensson TA, Nordenström J. Seventeen year follow-up study of bone mass in patients with mild asymptomatic hyperparathyroidism some of whom were operated on. Eur J Surg. 1995;161(12):863-9.
39. Paterson CR, Burns J, Mowat E. Long term follow up of untreated primary hyperparathyroidism. Br Med J (Clin Res Ed). 1984;289(6454):1261-3.
40. Parfitt AM. Equilibrium and disequilibrium hypercalcemia new light on an old concept. Metabolic Bone Disease and Related Research.1(4):279-93.
41. Bilezikian JP, Khan AA, Silverberg SJ, Fuleihan GE, Marcocci C, Minisola S, et al. Evaluation and Management of Primary Hyperparathyroidism: Summary Statement and Guidelines from the Fifth

International Workshop. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2022.
42. NIH conference. Diagnosis and management of asymptomatic primary hyperparathyroidism: consensus development conference statement. Annals of internal medicine. 1991;114(7):593-7.
43. Bilezikian JP, Potts JT, Jr., Fuleihan Gel H, Kleerekoper M, Neer R, Peacock M, et al. Summary statement from a workshop on asymptomatic primary hyperparathyroidism: a perspective for the 21st century. J Clin Endocrinol Metab. 2002;87(12):5353-61.
44. Bilezikian JP, Khan AA, Potts JT, Jr. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the third international workshop. J Clin Endocrinol Metab. 2009;94(2):335-9.
45. Bilezikian JP, Brandi ML, Eastell R, Silverberg SJ, Udelsman R, Marcocci C, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. J Clin Endocrinol Metab. 2014;99(10):3561-9.
46. Bollerslev J, Rejnmark L, Zahn A, Heck A, Appelman-Dijkstra NM, Cardoso L, et al. European Expert Consensus on Practical Management of Specific Aspects of Parathyroid Disorders in Adults and in Pregnancy: Recommendations of the ESE Educational Program of Parathyroid Disorders. European journal of endocrinology. 2022;186(2):R33-r63.
47. Insogna KL. Primary Hyperparathyroidism. The New England journal of medicine. 2018;379(25):e43.
48. Payne RB, Little AJ, Williams RB, Milner JR. Interpretation of serum calcium in patients with abnormal serum proteins. British medical journal. 1973;4(5893):643-6.
49. Messerli FH, Bangalore S. Half a century of hydrochlorothiazide: facts, fads, fiction, and follies. The American journal of medicine. 2011;124(10):896-9.
50. Griebeler ML, Kearns AE, Ryu E, Thapa P, Hathcock MA, Melton LJ, 3rd, et al. Thiazide-Associated Hypercalcemia: Incidence and Association With Primary Hyperparathyroidism Over Two Decades. J Clin Endocrinol Metab. 2016;101(3):1166-73.
51. Lajeunesse D, Bouhtiauy I, Brunette MG. Parathyroid hormone and hydrochlorothiazide increase calcium transport by the luminal membrane of rabbit distal nephron segments through different pathways. Endocrinology. 1994;134(1):35-41.
52. Yacobi-Bach M, Serebro M, Greenman Y, Tordjman K, Stern N. Letter to the editor: Thiazides are not inducers of PTH secretion: a comment on normocalcemic hyperparathyroidism. J Clin Endocrinol Metab. 2015;100(2):L27-8.
53. Meehan AD, Udumyan R, Kardell M, Landén M, Järhult J, Wallin G. Lithium-Associated Hypercalcemia: Pathophysiology, Prevalence, Management. World journal of surgery. 2018;42(2):415-24.
54. Shapiro HI, Davis KA. Hypercalcemia and "primary" hyperparathyroidism during lithium therapy. Am J Psychiatry. 2015;172(1):12-5.
55. Skandarajah AR, Palazzo FF, Henry JF. Lithium-associated hyperparathyroidism: surgical strategies in the era of minimally invasive parathyroidectomy. World journal of surgery. 2011;35(11):2432-9.
56. Schulte KM, Talat N. Diagnosis and management of parathyroid cancer. Nat Rev Endocrinol. 2012;8(10):612-22.
57. Rodrigo JP, Hernandez-Prera JC, Randolph GW, Zafereo ME, Hartl DM, Silver CE, et al. Parathyroid cancer: An update. Cancer Treat Rev. 2020;86:102012.
58. Palmer M, Adami HO, Bergström R, Jakobsson S, Akerström G, Ljunghall S. Survival and renal function in untreated hypercalcaemia. Population-based cohort study with 14 years of follow-up. Lancet (London, England). 1987;1(8524):59-62.
59. Øgard CG, Engholm G, Almdal TP, Vestergaard H. Increased mortality in patients hospitalized with primary hyperparathyroidism during the period 1977-1993 in Denmark. World journal of surgery. 2004;28(1):108-11.
60. Clifton-Bligh PB, Nery ML, Supramaniam R, Reeve TS, Delbridge L, Stiel JN, et al. Mortality associated with primary hyperparathyroidism. Bone. 2015;74:121-4.

61. Jorde R, Sundsfjord J, Fitzgerald P, Bonaa KH. Serum calcium and cardiovascular risk factors and diseases: the Tromso study. Hypertension (Dallas, Tex : 1979). 1999;34(3):484-90.
62. Wermers RA, Khosla S, Atkinson EJ, Grant CS, Hodgson SF, O'Fallon WM, et al. Survival after the diagnosis of hyperparathyroidism: a population-based study. The American journal of medicine. 1998;104(2):115-22.
63. Nilsson IL, Zedenius J, Yin L, Ekbom A. The association between primary hyperparathyroidism and malignancy: nationwide cohort analysis on cancer incidence after parathyroidectomy. Endocrine-related cancer. 2007;14(1):135-40.
64. McMahon DJ, Carrelli A, Palmeri N, Zhang C, DiTullio M, Silverberg SJ, et al. Effect of Parathyroidectomy Upon Left Ventricular Mass in Primary Hyperparathyroidism: A Meta-Analysis. J Clin Endocrinol Metab. 2015;100(12):4399-407.
65. Persson A, Bollerslev J, Rosen T, Mollerup CL, Franco C, Isaksen GA, et al. Effect of surgery on cardiac structure and function in mild primary hyperparathyroidism. Clinical endocrinology. 2011;74(2):174-80.
66. Schillaci G, Pucci G, Pirro M, Monacelli M, Scarponi AM, Manfredelli MR, et al. Large-artery stiffness: a reversible marker of cardiovascular risk in primary hyperparathyroidism. Atherosclerosis. 2011;218(1):96-101.
67. Bolland MJ, Grey AB, Gamble GD, Reid IR. Association between primary hyperparathyroidism and increased body weight: a meta-analysis. J Clin Endocrinol Metab. 2005;90(3):1525-30.
68. Pepe J, Cipriani C, Sonato C, Raimo O, Biamonte F, Minisola S. Cardiovascular manifestations of primary hyperparathyroidism: a narrative review. European journal of endocrinology. 2017;177(6):R297-r308.
69. Godang K, Lundstam K, Mollerup C, Fougner F, Pernow Y, Nordenström J, et al. The effect of surgery on fat mass, lipid and glucose metabolism in mild primary hyperparathyroidism. Endocrine connections. 2018;7(8):941-8.
70. Ambrogini E, Cetani F, Cianferotti L, Vignali E, Banti C, Viccica G, et al. Surgery or surveillance for mild asymptomatic primary hyperparathyroidism: a prospective, randomized clinical trial. J Clin Endocrinol Metab. 2007;92(8):3114-21.
71. Rao DS, Phillips ER, Divine GW, Talpos GB. Randomized controlled clinical trial of surgery versus no surgery in patients with mild asymptomatic primary hyperparathyroidism. J Clin Endocrinol Metab. 2004;89(11):5415-22.
72. Ejlsmark-Svensson H, Rolighed L, Rejnmark L. Effect of Parathyroidectomy on Cardiovascular Risk Factors in Primary Hyperparathyroidism: A Randomized Clinical Trial. The Journal of Clinical Endocrinology & Metabolism. 2019;104(8):3223-32.
73. WHO. The World Health Organization Quality of Life (WHOQOL) 2012 [Available from: https://www.who.int/toolkits/whoqol.
74. Giudice ML, Mihalik B, Dinnyés A, Kobolák J. The Nervous System Relevance of the Calcium Sensing Receptor in Health and Disease. Molecules. 2019;24(14).
75. Cheng SP, Lee JJ, Liu TP, Yang PS, Liu SC, Hsu YC, et al. Quality of Life After Surgery or Surveillance for Asymptomatic Primary Hyperparathyroidism: A Meta-Analysis of Randomized Controlled Trials. Medicine. 2015;94(23):e931.
76. Bollerslev J, Jansson S, Mollerup CL, Nordenstrom J, Lundgren E, Torring O, et al. Medical observation, compared with parathyroidectomy, for asymptomatic primary hyperparathyroidism: a prospective, randomized trial. J Clin Endocrinol Metab. 2007;92(5):1687-92.
77. Horiuchi K, Yoshida Y, Okamoto T. Effects of surgery on the patient-reported outcomes of primary hyperparathyroidism patients with mild hypercalcemia without classic symptoms: a systematic review of the literature. Surgery today. 2019.
78. Udelsman R, Åkerström G, Biagini C, Duh QY, Miccoli P, Niederle B, et al. The surgical management of asymptomatic primary hyperparathyroidism: proceedings of the Fourth International Workshop. J Clin Endocrinol Metab. 2014;99(10):3595-606.

79. Perrier N, Lang BH, Farias LCB, Poch LL, Sywak M, Almquist M, et al. Surgical Aspects of Primary Hyperparathyroidism. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2022;37(11):2373-90.
80. Shawky M, Abdel Aziz T, Morley S, Beale T, Bomanji J, Soromani C, et al. Impact of intraoperative parathyroid hormone monitoring on the management of patients with primary hyperparathyroidism. Clinical endocrinology. 2019;90(2):277-84.
81. Chen H, Pruhs Z, Starling JR, Mack E. Intraoperative parathyroid hormone testing improves cure rates in patients undergoing minimally invasive parathyroidectomy. Surgery. 2005;138(4):583-7; discussion 7-90.
82. Marcocci C, Bollerslev J, Khan AA, Shoback DM. Medical management of primary hyperparathyroidism: proceedings of the fourth International Workshop on the Management of Asymptomatic Primary Hyperparathyroidism. J Clin Endocrinol Metab. 2014;99(10):3607-18.
83. Papapoulos SE. Bisphosphonates: how do they work? Best practice & research Clinical endocrinology & metabolism. 2008;22(5):831-47.
84. Leere JS, Karmisholt J, Robaczyk M, Vestergaard P. Contemporary Medical Management of Primary Hyperparathyroidism: A Systematic Review. Frontiers in endocrinology. 2017;8:79.
85. Orr LE, Zhou H, Zhu CY, Haigh PI, Adams AL, Yeh MW. Skeletal effects of combined medical and surgical management of primary hyperparathyroidism. Surgery. 2020;167(1):144-8.
86. Ryhänen EM, Koski AM, Löyttyniemi E, Välimäki MJ, Kiviniemi U, Schalin-Jäntti C. Postoperative zoledronic acid for osteoporosis in primary hyperparathyroidism: a randomized placebo-controlled study. European journal of endocrinology. 2021;185(4):515-24.
87. Corsello SM, Paragliola RM, Locantore P, Ingraudo F, Ricciato MP, Rota CA, et al. Post-surgery severe hypocalcemia in primary hyperparathyroidism preoperatively treated with zoledronic acid. Hormones (Athens). 2010;9(4):338-42.
88. Dandurand K, Ali DS, Khan AA. Primary Hyperparathyroidism: A Narrative Review of Diagnosis and Medical Management. J Clin Med. 2021;10(8).
89. Nagano N. Pharmacological and clinical properties of calcimimetics: calcium receptor activators that afford an innovative approach to controlling hyperparathyroidism. Pharmacol Ther. 2006;109(3):339-65.
90. Valle C, Rodriguez M, Santamaría R, Almaden Y, Rodriguez ME, Cañadillas S, et al. Cinacalcet reduces the set point of the PTH-calcium curve. J Am Soc Nephrol. 2008;19(12):2430-6.
91. Ng CH, Chin YH, Tan MHQ, Ng JX, Yang SP, Kiew JJ, et al. Cinacalcet and primary hyperparathyroidism: systematic review and meta regression. Endocrine connections. 2020;9(7):724-35.
92. Faggiano A, Di Somma C, Ramundo V, Severino R, Vuolo L, Coppola A, et al. Cinacalcet hydrochloride in combination with alendronate normalizes hypercalcemia and improves bone mineral density in patients with primary hyperparathyroidism. Endocrine. 2011;39(3):283-7.
93. Delmas PD. Clinical potential of RANKL inhibition for the management of postmenopausal osteoporosis and other metabolic bone diseases. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry. 2008;11(2):325-38.
94. Leere JS, Karmisholt J, Robaczyk M, Lykkeboe S, Handberg A, Steinkohl E, et al. Denosumab and cinacalcet for primary hyperparathyroidism (DENOCINA): a randomised, double-blind, placebo-controlled, phase 3 trial. The lancet Diabetes & endocrinology. 2020;8(5):407-17.
95. Eller-Vainicher C, Palmieri S, Cairoli E, Goggi G, Scillitani A, Arosio M, et al. Protective Effect of Denosumab on Bone in Older Women with Primary Hyperparathyroidism. J Am Geriatr Soc. 2018;66(3):518-24.
96. Cummings SR, San Martin J, McClung MR, Siris ES, Eastell R, Reid IR, et al. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. The New England journal of medicine. 2009;361(8):756-65.
97. Wagner EH, Bledsoe T. The Rand Health Insurance Experiment and HMOs. Medical care. 1990;28(3):191-200.

98. Tarlov AR, Ware JE, Jr, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study: An Application of Methods for Monitoring the Results of Medical Care. Jama. 1989;262(7):925-30.
99. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical care. 1992;30(6):473-83.
100. Ware JE. SF-36 Health Survey- Manual and Interpretation Guide: Quality Metric Incorporated; 2005.
101. Jenkinson C, Layte R, Jenkinson D, Lawrence K, Petersen S, Paice C, et al. A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? J Public Health Med. 1997;19(2):179-86.
102. Gandek B, Ware JE, Jr. Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. International Quality of Life Assessment. Journal of clinical epidemiology. 1998;51(11):953-9.
103. Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. Journal of clinical epidemiology. 1998;51(11):903-12.
104. Ware JE, Snow, K.K., Kosinski, M., et al. , . SF-36 Health Survey Manual and Interpretation Guide. . New England Medical Center: The Health Institute, Boston.; 1993.
105. PROMiNET. Welcome to PROMiNET https://www.prominet.no/index.php?page=english2022 [
106. Enden T, Bernklev T, Jelsness-Jørgensen LP, Amdal CD. [Patients know their own health best]. Tidsskrift for den Norske laegeforening : tidsskrift for praktisk medicin, ny raekke. 2018;138(8).
107. Asberg M, Montgomery SA, Perris C, Schalling D, Sedvall G. A comprehensive psychopathological rating scale. Acta Psychiatr Scand Suppl. 1978(271):5-27.
108. Svanborg P, Asberg M. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. Acta Psychiatr Scand. 1994;89(1):21-8.
109. Lundgren E, Ljunghall S, Akerstrom G, Hetta J, Mallmin H, Rastad J. Case-control study on symptoms and signs of "asymptomatic" primary hyperparathyroidism. Surgery. 1998;124(6):980-5; discussion 5-6.
110. Pretorius M, Lundstam K, Heck A, Fagerland MW, Godang K, Mollerup C, et al. Mortality and Morbidity in Mild Primary Hyperparathyroidism: Results From a 10-Year Prospective Randomized Controlled Trial of Parathyroidectomy Versus Observation. Annals of internal medicine. 2022;175(6):812-9.
111. Gluer CC. 30years of DXA technology innovations. Bone. 2017;104:7-12.
112. Genant HK, Engelke K, Fuerst T, Glüer CC, Grampp S, Harris ST, et al. Noninvasive assessment of bone mineral and structure: state of the art. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 1996;11(6):707-30.
113. Dimai HP. Use of dual-energy X-ray absorptiometry (DXA) for diagnosis and fracture risk assessment; WHO-criteria, T- and Z-score, and reference databases. Bone. 2017;104:39-43.
114. Kanis JA. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. WHO Study Group. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 1994;4(6):368-81.
115. Shuhart CR, Yeap SS, Anderson PA, Jankowski LG, Lewiecki EM, Morse LR, et al. Executive Summary of the 2019 ISCD Position Development Conference on Monitoring Treatment, DXA Cross-calibration and Least Significant Change, Spinal Cord Injury, Peri-prosthetic and Orthopedic Bone Health, Transgender Medicine, and Pediatrics. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry. 2019;22(4):453-71.
116. Laskey MA, Flaxman ME, Barber RW, Trafford S, Hayball MP, Lyttle KD, et al. Comparative performance in vitro and in vivo of Lunar DPX and Hologic QDR-1000 dual energy X-ray absorptiometers. Br J Radiol. 1991;64(767):1023-9.
117. Lundstam K, Godang K, Pretorius M, Markwardt P, Hellström M, Bollerslev J, et al. The Influence of Dxa Hardware, Software, Reference Population and Software Analysis Settings on the

Bone Mineral Density and T-Score Relationship. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry. 2021.
118. Akobeng AK. Understanding randomised controlled trials. Arch Dis Child. 2005;90(8):840-4.
119. Rush CJ, Campbell RT, Jhund PS, Petrie MC, McMurray JJV. Association is not causation: treatment effects cannot be estimated from observational data in heart failure. Eur Heart J. 2018;39(37):3417-38.
120. Lindekleiv HM, Due J, Svartberg J, Varhaug JE. [Treatment practice in primary hyperparathyroidism]. Tidsskrift for den Norske laegeforening : tidsskrift for praktisk medicin, ny raekke. 2009;129(4):300-2.
121. Kang M, Ragan BG, Park JH. Issues in outcomes research: an overview of randomization techniques for clinical trials. J Athl Train. 2008;43(2):215-21.
122. Pretorius M, Lundstam K, Hellstrom M, Fagerland MW, Godang K, Mollerup C, et al. Effects of Parathyroidectomy on Quality of Life: 10 Years of Data From a Prospective Randomized Controlled Trial on Primary Hyperparathyroidism (the SIPH-Study). J Bone Miner Res. 2020.
123. Lundstam K, Pretorius M, Bollerslev J, Godang K, Fagerland MW, Mollerup C, et al. Positive effect of parathyroidectomy compared to observation on BMD in a randomized controlled trial of mild primary hyperparathyroidism. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2023.
124. Torgerson DJ, Roberts C. Understanding controlled trials. Randomisation methods: concealment. BMJ (Clinical research ed). 1999;319(7206):375-6.
125. Bland JM, Altman DG. The logrank test. BMJ (Clinical research ed). 2004;328(7447):1073.
126. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis--an introduction to concepts and methods. Br J Cancer. 2003;89(3):431-6.
127. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society: Series B (Methodological). 1972;34(2):187-202.
128. Metzger SK, Jones BT. Properly calculating estat phtest in the presence of stratified hazards. The Stata Journal. 2021;21(4):1028-33.
129. Schober P, Vetter TR. Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. Anesth Analg. 2018;127(3):792-8.
130. Yang W, Jepson C, Xie D, Roy JA, Shou H, Hsu JY, et al. Statistical Methods for Recurrent Event Analysis in Cohort Studies of CKD. Clin J Am Soc Nephrol. 2017;12(12):2066-73.
131. Pekar JD, Grzych G, Durand G, Haas J, Lionet A, Brousseau T, et al. Calcium state estimation by total calcium: the evidence to end the never-ending story. Clin Chem Lab Med. 2020;58(2):222-31.
132. Lian IA, Åsberg A. Should total calcium be adjusted for albumin? A retrospective observational study of laboratory data from central Norway. BMJ Open. 2018;8(4):e017703.
133. Ong GS, Walsh JP, Stuckey BG, Brown SJ, Rossi E, Ng JL, et al. The importance of measuring ionized calcium in characterizing calcium status and diagnosing primary hyperparathyroidism. J Clin Endocrinol Metab. 2012;97(9):3138-45.
134. Slomp J, van der Voort PH, Gerritsen RT, Berk JA, Bakker AJ. Albumin-adjusted calcium is not suitable for diagnosis of hyper- and hypocalcemia in the critically ill. Crit Care Med. 2003;31(5):1389-93.
135. Walker MD, Silverberg SJ. Quality of Life in Primary Hyperparathyroidism Revisited: Keep Calm and Carry on? 2021;36(1):1-2.
136. Probst P, Grummich K, Harnoss JC, Hüttner FJ, Jensen K, Braun S, et al. Placebo-Controlled Trials in Surgery: A Systematic Review and Meta-Analysis. Medicine. 2016;95(17):e3516.
137. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990;1(1):43-6.
138. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled clinical trials. 1989;10(4):407-15.
139. Mouelhi Y, Jouve E, Castelli C, Gentile S. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. Health and quality of life outcomes. 2020;18(1):136.

140. Osoba D, Bezjak A, Brundage M, Zee B, Tu D, Pater J. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of The National Cancer Institute of Canada Clinical Trials Group. European journal of cancer (Oxford, England : 1990). 2005;41(2):280-7.
141. Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. BMJ (Clinical research ed). 1993;306(6890):1437-40.
142. Garratt AM, Stavem K. Measurement properties and normative data for the Norwegian SF-36: results from a general population survey. Health and quality of life outcomes. 2017;15(1):51.
143. Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? Qual Life Res. 2001;10(5):395-404.
144. Ware JE, Kosinski M. Interpreting SF-36 summary health measures: a response. Qual Life Res. 2001;10(5):405-13; discussion 15-20.
145. Ware JEDJEKM. How to score version 2 of the SF-36 health survey : (standard & acute forms) ; [SF-36v2]. Lincoln, RI: QualityMetric; 2001.
146. Hawthorne G, Osborne RH, Taylor A, Sansoni J. The SF36 Version 2: critical analyses of population weights, scoring algorithms and population norms. Qual Life Res. 2007;16(4):661-73.
147. Salinari G, De Santis G. On the beginning of mortality acceleration. Demography. 2015;52(1):39-60.
148. Phillips DE, Lozano R, Naghavi M, Atkinson C, Gonzalez-Medina D, Mikkelsen L, et al. A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. Population health metrics. 2014;12:14.
149. Pedersen AG, Ellingsen CL. Data quality in the Causes of Death Registry. Tidsskrift for den Norske laegeforening : tidsskrift for praktisk medicin, ny raekke. 2015;135(8):768-70.
150. Axelsson KF, Wallander M, Johansson H, Harvey NC, Vandenput L, McCloskey E, et al. Analysis of Comorbidities, Clinical Outcomes, and Parathyroidectomy in Adults With Primary Hyperparathyroidism. JAMA Netw Open. 2022;5(6):e2215396.
151. D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. Journal of clinical epidemiology. 1996;49(12):1429-33.
152. Drosdowsky A, Gough K. The Charlson Comorbidity Index: problems with use in epidemiological research. Journal of clinical epidemiology. 2022.
153. Banerjee A, Chitnis UB, Jadhav SL, Bhawalkar JS, Chaudhury S. Hypothesis testing, type I and type II errors. Ind Psychiatry J. 2009;18(2):127-31.
154. de Oliveira Martins Duarte J, Pereira PMLP, Lopes JNG, Goes ATL, Durán DC, Sobral ASGN, et al. A reversible case of rapidly progressive dementia—Hypercalcemia due to hyperparathyroidism. Clinical Case Reports. 2019;7(12):2571-4.
155. Singh H, Selvaraj V, Padala PR. Dementia secondary to hyperparathyroidism. Psychiatry (Edgmont). 2010;7(8):13-4.
156. Papageorgiou SG, Christou Y, Kontaxis T, Bonakis A, Anagnostouli M, Potagas C, et al. Dementia as presenting symptom of primary hyperparathyroidism: favourable outcome after surgery. Clin Neurol Neurosurg. 2008;110(10):1038-40.
157. Bolland MJG, A. Nonoperative Management of Mild Primary Hyperparathyroidism: A Reasonable, Evidence-Based Option. Annals of internal medicine. 2022;175(6):899-900.
158. Rejnmark L, Ejlsmark-Svensson H. Effects of PTH and PTH Hypersecretion on Bone: a Clinical Perspective. Current osteoporosis reports. 2020;18(3):103-14.
159. Dempster DW, Parisien M, Silverberg SJ, Liang XG, Schnitzer M, Shen V, et al. On the mechanism of cancellous bone preservation in postmenopausal women with mild primary hyperparathyroidism. J Clin Endocrinol Metab. 1999;84(5):1562-6.
160. Ott SM. Cortical or Trabecular Bone: What's the Difference? Am J Nephrol. 2018;47(6):373-5.
161. Dempster DW, Müller R, Zhou H, Kohler T, Shane E, Parisien M, et al. Preserved three-dimensional cancellous bone structure in mild primary hyperparathyroidism. Bone. 2007;41(1):19-24.
162. Goltzman D. Physiology of Parathyroid Hormone. Endocrinology and metabolism clinics of North America. 2018;47(4):743-58.

163. Wood K, Dhital S, Chen H, Sippel RS. What is the utility of distal forearm DXA in primary hyperparathyroidism? Oncologist. 2012;17(3):322-5.
164. Silverberg SJ, Shane E, Jacobs TP, Siris E, Bilezikian JP. A 10-year prospective study of primary hyperparathyroidism with or without parathyroid surgery. The New England journal of medicine. 1999;341(17):1249-55.
165. Rubin MR, Cosman F, Lindsay R, Bilezikian JP. The anabolic effects of parathyroid hormone. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2002;13(4):267-77.
166. Riggs BL, Wahner HW, Dunn WL, Mazess RB, Offord KP, Melton LJ, 3rd. Differential changes in bone mineral density of the appendicular and axial skeleton with aging: relationship to spinal osteoporosis. The Journal of clinical investigation. 1981;67(2):328-35.
167. Parisien M, Silverberg SJ, Shane E, Dempster DW, Bilezikian JP. Bone disease in primary hyperparathyroidism. Endocrinology and metabolism clinics of North America. 1990;19(1):19-34.
168. El-Hajj Fuleihan G, Chakhtoura M, Cipriani C, Eastell R, Karonova T, Liu JM, et al. Classical and Nonclassical Manifestations of Primary Hyperparathyroidism. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2022.
169. Ejlsmark-Svensson H, Rolighed L, Harsløf T, Rejnmark L. Risk of fractures in primary hyperparathyroidism: a systematic review and meta-analysis. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2021;32(6):1053-60.
170. Vestergaard P, Mollerup CL, Frøkjaer VG, Christiansen P, Blichert-Toft M, Mosekilde L. Cohort study of risk of fracture before and after surgery for primary hyperparathyroidism. BMJ (Clinical research ed). 2000;321(7261):598-602.
171. De Geronimo S, Romagnoli E, Diacinti D, D'Erasmo E, Minisola S. The risk of fractures in postmenopausal women with primary hyperparathyroidism. European journal of endocrinology. 2006;155(3):415-20.
172. Harvey NC, Glüer CC, Binkley N, McCloskey EV, Brandi ML, Cooper C, et al. Trabecular bone score (TBS) as a new complementary approach for osteoporosis evaluation in clinical practice. Bone. 2015;78:216-24.
173. Grigorie D, Coles D, Sucaliuc A. TRABECULAR BONE SCORE (TBS) HAS A POOR DISCRIMINATIVE POWER FOR VERTEBRAL FRACTURES IN 153 ROMANIAN PATIENTS WITH PRIMARY HYPERPARATHYROIDISM. Acta Endocrinol (Buchar). 2018;14(2):208-12.
174. Liu M, Williams J, Kuo J, Lee JA, Silverberg SJ, Walker MD. Risk factors for vertebral fracture in primary hyperparathyroidism. Endocrine. 2019;66(3):682-90.
175. Mikolajewicz N, Bishop N, Burghardt AJ, Folkestad L, Hall A, Kozloff KM, et al. HR-pQCT Measures of Bone Microarchitecture Predict Fracture: Systematic Review and Meta-Analysis. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2020;35(3):446-59.
176. Stein EM, Silva BC, Boutroy S, Zhou B, Wang J, Udesky J, et al. Primary hyperparathyroidism is associated with abnormal cortical and trabecular microstructure and reduced bone stiffness in postmenopausal women. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. 2013;28(5):1029-40.
177. Wang W, Nie M, Jiang Y, Li M, Meng X, Xing X, et al. Impaired geometry, volumetric density, and microstructure of cortical and trabecular bone assessed by HR-pQCT in both sporadic and MEN1-related primary hyperparathyroidism. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2020;31(1):165-73.
178. Schoeb M, Hamdy NAT, Malgo F, Winter EM, Appelman-Dijkstra NM. Added Value of Impact Microindentation in the Evaluation of Bone Fragility: A Systematic Review of the Literature. Frontiers in endocrinology. 2020;11:15.

179. Schoeb M, Winter EM, Sleddering MA, Lips MA, Schepers A, Snel M, et al. Bone Material Strength Index as Measured by Impact Microindentation is Low in Patients with Primary Hyperparathyroidism. J Clin Endocrinol Metab. 2021;106(7):e2527-e34.
180. Fan B, Lu Y, Genant H, Fuerst T, Shepherd J. Does standardized BMD still remove differences between Hologic and GE-Lunar state-of-the-art DXA systems? Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2010;21(7):1227-36.
181. Morrison SA, Petri RM, Hunter HL, Raju D, Gower B. Comparison of the Lunar Prodigy and iDXA Dual-Energy X-ray Absorptiometers for Assessing Total and Regional Body Composition. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry. 2016;19(3):290-7.
182. Huffman DM, Landy NM, Potter E, Nagy TR, Gower BA. Comparison of the Lunar DPX-L and Prodigy dual-energy X-ray absorptiometers for assessing total and regional body composition. Int J Body Compos Res. 2005;3(1):25-30.
183. Emaus N, Omsland TK, Ahmed LA, Grimnes G, Sneve M, Berntsen GK. Bone mineral density at the hip in Norwegian women and men--prevalence of osteoporosis depends on chosen references: the Tromsø Study. Eur J Epidemiol. 2009;24(6):321-8.
184. Lundstam K, Godang K, Pretorius M, Markwardt P, Hellström M, Bollerslev J, et al. The Influence of DXA Hardware, Software, Reference Population and Software Analysis Settings on the Bone Mineral Density and T-Score Relationship. Journal of clinical densitometry : the official journal of the International Society for Clinical Densitometry. 2022;25(1):24-33.
185. ISCD. ISCD- Official-Positions-Adult-1.pdf 2019 [Available from: https://iscd.org/wp-content/uploads/2021/09/2019-Official-Positions-Adult-1.pdf.
186. Shepherd JA, Ng BK, Sommer MJ, Heymsfield SB. Body composition by DXA. Bone. 2017;104:101-5.

I

CLINICAL TRIAL

JBMR®

Effects of Parathyroidectomy on Quality of Life: 10 Years of Data From a Prospective Randomized Controlled Trial on Primary Hyperparathyroidism (the SIPH-Study)

Mikkel Pretorius,[1,2] Karolina Lundstam,[3,4] Mikael Hellström,[3,4] Morten W Fagerland,[5] Kristin Godang,[1] Charlotte Mollerup,[6] Stine L Fougner,[7] Ylva Pernow,[8] Turid Aas,[9] Ola Hessman,[10] Thord Rosén,[11] Jörgen Nordenström,[12] Svante Jansson,[13] Ansgar Heck,[1,2†] and Jens Bollerslev[1,2†]

[1]Section of Specialized Endocrinology, Oslo University Hospital, Oslo, Norway
[2]Faculty of Medicine, University of Oslo, Oslo, Norway
[3]Department of Radiology, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[4]Department of Radiology, Sahlgrenska University Hospital, Gothenburg, Sweden
[5]Oslo Centre for Biostatistics and Epidemiology, Research Support Services, Oslo University Hospital, Oslo, Norway
[6]Clinic of Breast and Endocrine Surgery, HovedOrtoCentret (HOC), Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
[7]Department of Endocrinology, Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
[8]Department of Molecular Medicine and Surgery, Department of Endocrinology, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
[9]Department of Breast and Endocrine Surgery, Haukeland University Hospital, Bergen, Norway
[10]Department of Surgery, Västmanland County Hospital, Västerås, Sweden
[11]Department of Medicine, Section of Endocrinology, Diabetes, and Metabolism, Sahlgrenska University Hospital, Gothenburg, Sweden
[12]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
[13]Department of Surgery, Institute of Clinical Sciences, Sahlgrenska Academy at the University of Gothenburg, Sahlgrenska University Hospital, Gothenburg, Sweden

ABSTRACT

Primary hyperparathyroidism (PHPT) was previously considered a disease presenting with multiorgan involvement and a wide range of symptoms. Today, the disease presents with no symptoms or mild symptomatology in most patients. Data regarding nonspecific symptoms such as pain, fatigue, memory loss, depression, and other neuropsychiatric signs have been ambiguous, and results from prospective long-term randomized control trials are lacking. The Scandinavian Investigation on Primary Hyperparathyroidism (SIPH) is a prospective randomized controlled trial (RCT) with 10-year follow up, comparing parathyroidectomy (PTX) to observation without any treatment (OBS). From 1998 to 2005, 191 patients with mild PHPT were included from Sweden, Norway, and Denmark. A total of 95 patients were randomized to PTX and 96 to OBS. The generic Short Form-36 survey (SF-36) and the Comprehensive Psychopathological Rating Scale (CPRS) were studied at baseline, 2, 5, and 10 years after randomization. After 10 years, the PTX group scored significantly better on vitality (PTX 65.1 ± 20.2 versus OBS 57.4 ± 22.7; $p = .017$) compared to the OBS group in SF-36. We found no differences between the groups in the physical subscales. The OBS group had no significant change in any of the SF-36 scores throughout the study. The CPRS showed an improvement of symptoms in both groups for single items and sum scores after 10 years compared to baseline. There were, however, no significant differences between the two groups in the CPRS data. The results of this large and long-term RCT indicate improvement in some of the mental domains of SF-36 following PTX. However, the treatment effects between the groups were subtle with uncertain clinical significance. The observation group had stable SF-36 values and improvement in CPRS symptom-scores. Thus, in considering only quality of life (QoL) and in the absence of declines in renal and skeletal parameters, it may be safe to observe patients with mild PHPT for a decade. © 2020 American Society for Bone and Mineral Research (ASBMR).

KEY WORDS: PARATHYROID-RELATED DISORDERS; CLINICAL TRIALS; PTH; HYPERCALCEMIA; RANDOMIZED CONTROLLED TRIAL

Received in original form May 27, 2020; revised form September 28, 2020; accepted October 7, 2020; Accepted manuscript online October 30, 2020.
Address correspondence to: Mikkel Pretorius, MD, Section of Specialized Endocrinology, Oslo University Hospital, Oslo, Norway. E-mail: mikkel.pretorius@ous-hf.no
Additional Supporting Information may be found in the online version of this article.
[†]AH and JB are joint final authors.
Journal of Bone and Mineral Research, Vol. 36, No. 1, January 2021, pp 3–11.
DOI: 10.1002/jbmr.4199

Introduction

Primary hyperparathyroidism (PHPT) is a common endocrine disorder among the elderly, especially among women.[1] Nowadays, most patients present with subtle or nonspecific symptoms in the developed societies,[2] but classic symptoms with multiorgan involvement and a wide range of symptoms are still present in some areas.[3,4] The change in clinical presentation is mainly due to increasing availability of routine analyses of serum calcium levels.[5,6] Parathyroidectomy (PTX) is the only curative treatment of PHPT with a high success rate,[1] and should be offered to patients with clear indications, such as osteoporosis or kidney stones.[2,7] However, because most patients do not present with such indications, observation with regular follow-up is recommended to most patients with the common, mild form of PHPT.[2,7]

Patients with chronic endocrine diseases, such as PHPT, have decreased quality of life (QoL). This is seen also in patients with mild PHPT, and positive effects of curative surgery for PHPT have been found in observational studies.[8–10] However, in short-term randomized controlled trials (RCTs), the results have been unclear with ambiguous findings.[11–13] Most studies were based on generic questionnaires, because disease-specific questionnaires were developed and validated only recently.[14] The disease-specific primary hyperparathyroidism quality-of-life questionnaire (PHPQoL) seems reliable, and correlates well with generic surveys and is sensitive to calcium related symptoms.[14] So far, the PHPQoL has not been tested in randomized studies, but a treatment effect of PTX was recently demonstrated in an observational study.[15] Due to the ambiguous results, symptoms related to cognition and neuropsychiatric manifestations have hitherto not been an indication for treatment.[2,7] However, long-term QoL results from well-designed controlled RCTs might have an impact on the guidance of patients with mild PHPT.

Disease activity is remarkably stable in mild PHPT,[16–18] in accordance with a set point error in the calcium-PTH regulation.[19] In mild PHPT, decision on management refers to calcium levels rather than PTH, the metabolic active hormone.[2,7] Calcium fluxes easily in different tissues, whereas PTH binds to the specific receptors PTHR1 and PTHR2. Both receptors are widely expressed in the body, including in the brain.[20,21] The distribution of PTHR1 and PTHR2 in the human brain indicates involvement in the regulation of fear, anxiety, the release of pituitary hormones, and nociception.[22] Thus, central nervous system (CNS)-related symptoms might reflect PTH as well as calcium levels, and potentially both.[23]

The Scandinavian Study on Primary Hyperparathyroidism (SIPH) was developed in the 1990s based on the NIH-initiated, first international workshop on management of asymptomatic (mild) PHPT[24] with the first patient included in 1998. QoL results, based on 2-year data, did not demonstrate a clear benefit of surgery compared with observation without intervention, in the SIPH study.[12] Ten years of follow-up concluded in 2016, and we here present the end of study results on QoL and neuropsychiatric symptoms from the hitherto largest and so far longest lasting RCT on the topic.

The aim of this study was to investigate the long-term effect of PTX compared with observation without intervention, in patients with mild PHPT. Moreover, we aimed to investigate if changes in calcium or PTH levels had an impact on QoL over time.

Patients and Methods

Participants

The SIPH trial (clinical trials.gov: NCT00522028) was a prospective randomized controlled trial with 10-year follow-up. The 191 enrolled patients (26 males), were block randomized to PTX ($n = 95$) or observation (OBS, $n = 96$). Patients were included from 1998 to 2005 in Sweden, Norway and Denmark.

The focus of the study was on patients with mild (previously termed asymptomatic) PHPT in accordance with the 1990 NIH criteria[24] for asymptomatic PHPT. Key inclusion criteria were untreated disease, with serum calcium (s-calcium) values between 2.60 and 2.80 mmol/L and age between 50 and 80 years. Table 1 presents baseline values from the patients available for QoL analyses. The mean ± SD age was 62.9 ± 7.4 years and mean ± SD s-calcium value was 2.65 mmol/L ± 0.11 at inclusion.

Table 1. Baseline Characteristics of the Patients Available for QoL Analysis

Characteristic	PTX ($n = 89$)	OBS ($n = 90$)
Age (years), mean ± SD	62.7 ± 7.9	63.1 ± 7.4
Gender, *n* (%)		
Female	76 (85.4)	77 (85.6)
Male	13 (14.6)	13 (14.4)
s-Calcium (mmol/L), mean ± SD[a]	2.65 ± 0.11	2.64 ± 0.11
s-Calcium (mg/dL), mean ± SD[a]	10.62 ± 0.44	10.58 ± 0.44
p-PTH (pmol/L), mean ± SD	10.16 ± 4.08	10.55 ± 4.26
s-Creatinine (μmol/L), mean ± SD	81.4 ± 17.4	79.7 ± 13.90
SF-36 domains, mean ± SD		
Physical functioning (PF)	73.82 ± 23.39	77.43 ± 20.75
Role physical (RP)	64.32 ± 41.09	67.97 ± 40.41
Bodily pain (BP)	69.66 ± 29.07	69.95 ± 28.27
General health (GH)	64.19 ± 22.70	66.74 ± 20.60
Vitality (VT)	54.27 ± 26.18	60.06 ± 26.61
Social functioning (SF)	81.68 ± 23.74	83.88 ± 22.96
Role emotional (RE)	64.15 ± 40.22	71.91 ± 41.10
Mental health (MH)	74.23 ± 19.77	74.74 ± 23.04
CPRS sum score	20.17 ± 16.14	19.44 ± 16.68

Data are presented as mean ± SD unless otherwise stated in the table.
OBS = observation; PTX = parathyroidectomy; p-PTH = plasma parathyroid hormone; QoL = quality of life; s-Calcium = serum calcium; s-Creatinine = serum creatinine.
[a]s-Calcium values are albumin-corrected.

Biochemistry

The biochemical analyses presented here (Table 1) were all measured consecutively by accredited medical biochemistry laboratories according to standard protocols. Calcium levels were measured as serum total calcium and corrected for variation in albumin by using the following formula: calcium adjusted for albumin = [calcium] mmol/L − (0.02 × ([albumin] g/L − 40)). All centers measured intact PTH (second-generation PTH assay), and levels were converted to International System (SI) units, when necessary.

QoL

Here, we focus on QoL throughout the 10-year study period. The generic Short Form 36 survey version 1 (SF-36) and the Comprehensive Psychopathological Rating Scale (CPRS) were performed

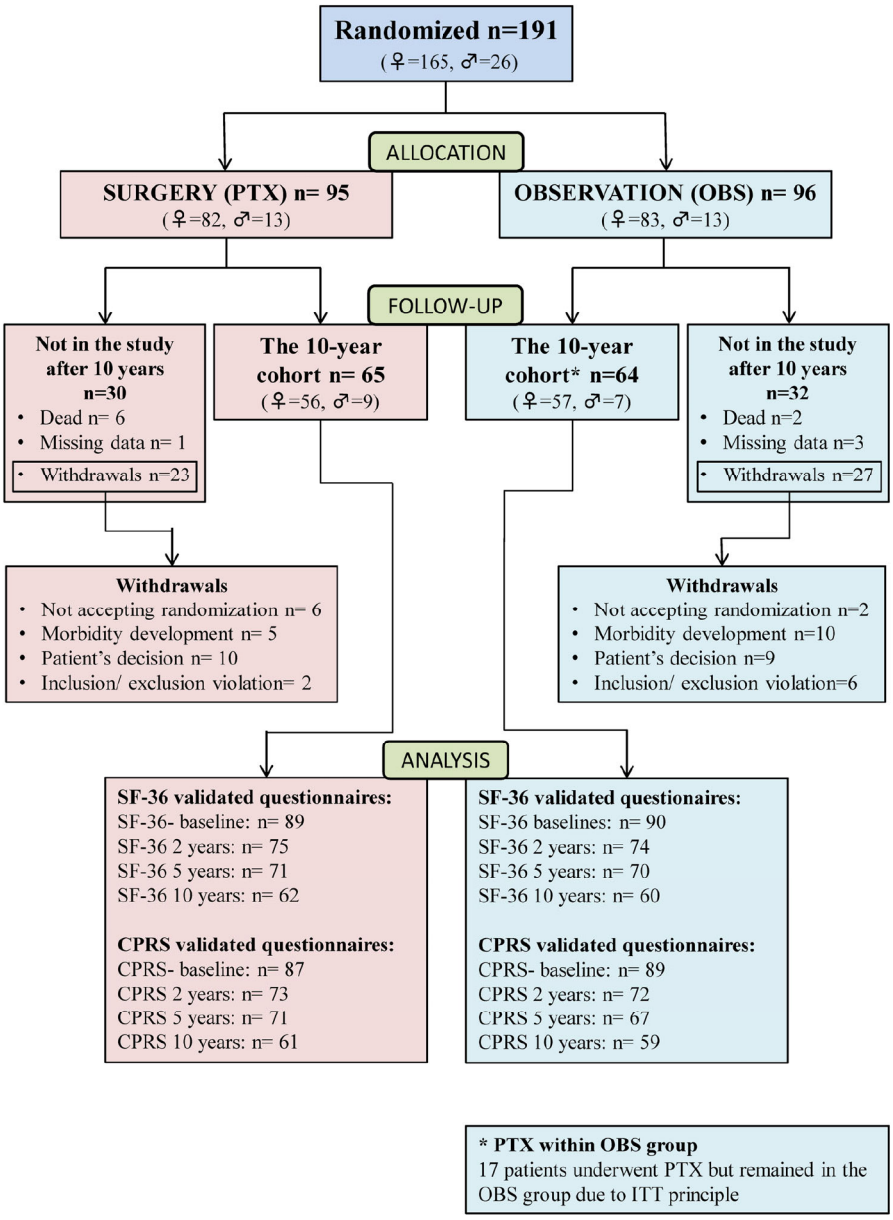

Fig 1. CONSORT flow diagram for study participants in the SIPH-trial. CONSORT = Consolidated Standards of Reporting Trials.

per protocol at baseline, 2, 5, and 10 years following randomization. The primary endpoint in the SIPH study was mortality. QoL was one of two predefined secondary endpoints, the other being morbidity.

At the end of study (10 years), 129 patients remained in the study. Enrollment, follow-up, and available data for analyses are presented in the Consolidated Standards of Reporting Trials (CONSORT) flowchart, see Fig. 1. Eight patients had died, 50 withdrew, and data were missing for four patients. Figure 1 summarizes the reasons for withdrawals. Fifteen patients withdrew from the study because of developing morbidity. They withdrew because of dementia (six patients, five in the OBS group), cancer (four patients, two in the OBS group), cardiovascular disease (three patients, one in the OBS group), Parkinson's disease (one patient in the OBS group), and vertigo (one patient in the PTX group). During the course of the study, 17 patients in the OBS group underwent PTX for various reasons. In the statistical analyses, these crossover patients followed the intention-to-treat principle and they were thus regarded as belonging to the OBS group.

All patients gave informed consent prior to inclusion and the study was approved by the local ethical committees in all three countries. The study followed the Declaration of Helsinki II.

SF-36

The SF-36 questionnaire has 35 items that measure eight generic dimensions of health. Four of these dimensions are mainly

Table 2. SF-36-Results From Baseline to 10 Years With Treatment Effect of Parathyroidectomy Derived From Mixed Model Statistics

SF-36 Domains	Baseline mean (95% CI)	2 Years mean (95% CI)	5 Years mean (95% CI)	10 Years mean (95% CI)	Change from baseline to 10 years mean (95% CI)	Between-group difference in changes from baseline to 10 years mean (95% CI); *p*-value[a]
Physical functioning						0.30 (−5.19 to 5.78); $p = .92$
Observation	75.9 (71.8 to 80.1)	75.2 (71.2 to 79.2)	74.0 (70.0 to 78.1)	72.1 (67.3 to 77.0)	−3.80 (−7.77 to 0.17)	
Parathyroidectomy	76.0 (71.8 to 80.2)	75.3 (71.2 to 79.4)	74.2 (70.2 to 78.3)	72.5 (67.8 to 77.2)	−3.50 (−7.29 to 0.29)	
Role physical						−3.48 (−16.3 to 9.34); $p = .60$
Observation	64.7 (57.4 to 72.0)	65.6 (58.8 to 72.3)	66.8 (60.0 to 73.7)	69.0 (59.8 to 78.2)	4.25 (−5.05 to 13.55)	
Parathyroidectomy	62.8 (55.4 to 70.2)	62.9 (56.1 to 69.8)	63.2 (56.3 to 70.0)	63.6 (54.8 to 72.3)	0.77 (−8.06 to 9.60)	
Bodily pain						−2.41 (−9.81 to 5.00); $p = .52$
Observation	68.3 (63.1 to 73.4)	68.1 (63.2 to 73.0)	67.8 (62.8 to 72.7)	67.3 (61.2 to 73.4)	−0.96 (−6.29 to 4.37)	
Parathyroidectomy	69.0 (63.7 to 74.2)	68.3 (63.3 to 73.3)	67.3 (62.3 to 72.2)	65.6 (59.7 to 71.5)	−3.37 (−8.50 to 1.77)	
General health						1.37 (−3.94 to 6.67); $p = .61$
Observation	66.4 (62.2 to 70.6)	66.6 (62.6 to 70.7)	67.0 (62.9 to 71.1)	67.6 (62.8 to 72.4)	1.23 (−2.60 to 5.07)	
Parathyroidectomy	66.1 (61.8 to 70.3)	66.6 (62.5 to 70.7)	67.4 (63.3 to 71.4)	68.7 (64.0 to 73.3)	2.60 (−1.06 to 6.26)	
Vitality						8.00 (1.42 to 14.6); **$p = .017$**
Observation	60.1 (55.6 to 64.7)	59.8 (55.5 to 64.2)	59.4 (55.0 to 63.8)	58.6 (53.2 to 64.1)	−1.50 (−6.29 to 3.28)	
Parathyroidectomy	57.3 (52.6 to 62.0)	58.6 (54.1 to 63.0)	60.5 (56.1 to 65.0)	63.8 (58.5 to 69.1)	6.50 (1.97 to 11.03)[b]	
Social functioning						6.45 (−0.34 to 13.2); $p = .063$
Observation	82.3 (78.2 to 86.4)	82.1 (78.3 to 85.9)	81.8 (77.9 to 85.7)	81.3 (76.2 to 86.3)	−1.04 (−5.93 to 3.85)	
Parathyroidectomy	81.1 (76.9 to 85.3)	82.2 (78.3 to 86.1)	83.8 (79.9 to 87.7)	86.5 (81.7 to 91.4)	5.41 (0.70 to 10.13)[c]	
Role emotional						4.89 (−8.56 to 18.33); $p = .47$
Observation	68.5 (61.4 to 75.5)	69.0 (62.6 to 75.4)	69.9 (63.4 to 76.5)	71.4 (62.2 to 80.5)	2.91 (−6.84 to 12.67)	
Parathyroidectomy	68.7 (61.5 to 75.9)	70.3 (63.7 to 76.8)	72.6 (66.1 to 79.1)	76.5 (67.8 to 85.2)	7.79 (−1.45 to 17.05)	
Mental health						3.41 (−2.21 to 9.03); $p = .23$
Observation	75.1 (71.4 to 78.9)	75.0 (71.5 to 78.5)	74.8 (71.2 to 78.4)	74.5 (70.0 to 79.0)	−0.67 (−4.74 to −4.74)	
Parathyroidectomy	75.5 (71.6 to 79.3)	76.0 (72.4 to 79.6)	76.8 (73.2 to 80.4)	78.2 (73.9 to 82.5)	2.73 (−1.12 to 6.60)	

[a]Positive values in between group differences denote a better SF-36 score in favor of parathyroidectomy.
[b]$p = .005$.
[c]$p = .024$.

physical (Physical Functioning [PF], Role Physical [RP], Bodily Pain [BP], and General Health [GH]). The four mental dimensions are Vitality (VT), Social Functioning (SF), Role Emotional (RE), and Mental Health (MH). A single question measures the change in health over time and does not contribute to the mentioned eight dimensions.

We used the SF-36 V1 (Ware-36). The questionnaire was translated to Swedish, Norwegian, and Danish as part of the International Quality of Life Assessment Project (IQOLA) to ensure correct translation and validation across borders.[25,26] All item scores were recoded, summed, and transformed to a scale from 0 (worst possible QoL score) to 100 (best). Baseline data have previously been described compared to normative Swedish data.[12]

CPRS

The original CPRS was developed almost 50 years ago[27] for evaluating global change in psychopathology. The 67 items covers a wide range of psychological symptomatology. It was constructed to be used either in full length or as a tool where subscales could be drawn for particular psychiatric syndromes, exemplified by Montgomery Aasberg Depression Rating Scores (MADRS).[28] Each item is scored on a Likert-type scale from 0 to 7 with zero as the best and seven as the worst scored QoL. This also allows for summative scales. Half steps were included to increase the sensitivity, as recommended in the original CPRS.[27]

The modified version of CPRS used in the present study was developed with focus on PHPT related symptoms.[29] This modified version has 17 items presented as Q1 to Q17 in Fig. 3*A*,*B* and in a supplemental file (Supplemental Table S1).

Statistics

We fitted a linear mixed model to all available data from the baseline, 1-year, 2-year, 5-year, and 10-year measurements for each dimension of SF-36 and CPRS (including the sum scores). All models included fixed effects for the treatment group, time, and treatment × time interaction, and a random intercept. Based on plots of the observed data, we modeled the time development as linear from baseline to 10 years. Following model fit, we estimated within-group mean values at baseline, 2, 5, and 10 years after baseline, mean change from baseline to 10 years, and between-group difference in changes from baseline to 10 years, with 95% confidence intervals. Data values from this mixed model are presented in Table 2. We also performed a test of the null hypothesis of no difference between the PTX and OBS groups in changes from baseline to 10 years.

Two sensitivity analyses were performed. First, we repeated the linear mixed models after excluding postsurgery data from the 17 patients in the OBS group who later underwent PTX during the follow-up period. Then, we fitted the linear mixed models with an additional fixed effect for albumin corrected s-calcium levels and PTH levels, thus obtaining results adjusted for both.

The statistical analyses were performed with Stata/SE 16.1 (StataCorp LLC, College Station, TX, USA). All confidence intervals and *p* values are two-sided.

Results

Baseline characteristics

The total cohort for the present analyses consisted of 179 patients, 89 randomized to PTX, and 90 to observation without intervention (Fig. 1, Table 1). The two groups were well matched for age, gender, s-calcium, plasma PTH (p-PTH), and serum creatinine (s-creatinine) levels. There was no difference in baseline values for SF-36 or CPRS between the groups.

Biochemistry

At baseline, the s-calcium levels in PTX and OBS-group were similar (2.65 ± 0.11 mmol/L and 2.64 ± 0.11 mmol/L). After 10 years,

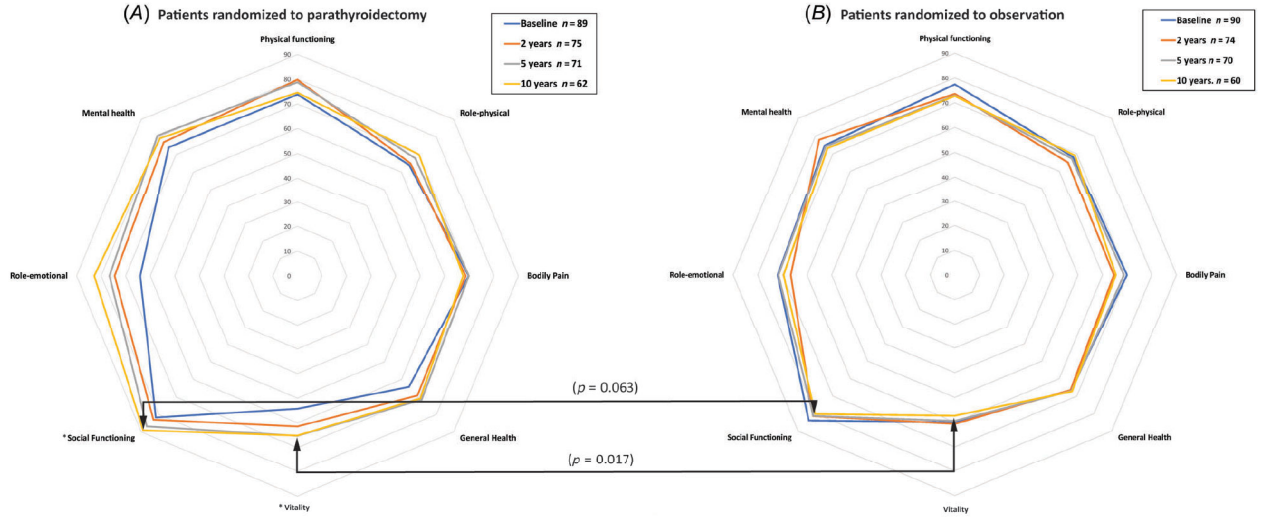

Fig 2. Radar chart of SF-36 mean values from baseline (blue line), 2 years (orange line), 5 years (gray line), and 10 years (yellow line) in PTX group (*A*) and OBS group (*B*). These descriptive data are calculated scores based directly on the SF-36 questionnaires. *Significant difference between baseline values and 10-year values within group. Values of *p* denote treatment effect between the groups over the 10-year period.

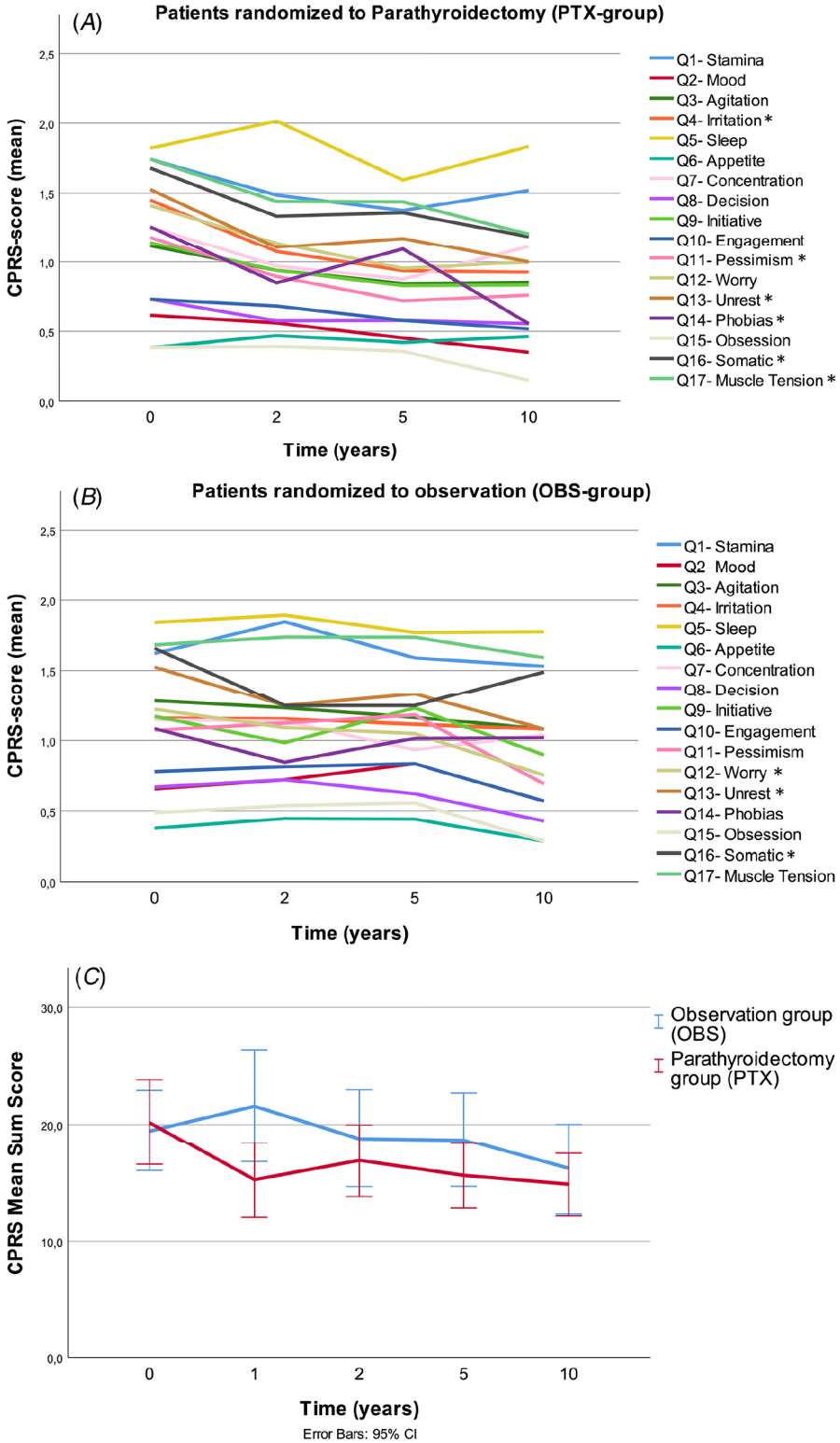

Fig 3. Mean CPRS scores for each item in PTX group (*A*) OBS group (*B*), and CPRS sumscores for both groups (*C*). *Significant difference between baseline values and 10-year values within the group. CPRS = Comprehensive Psychopathological Rating Scale.

the PTX group had a significant decrease in s-calcium to 2.35 mmol/L compared to baseline ($p < .001$). The OBS-group also declined in s-calcium levels to 2.53 ± 0.20 mmol/L after 10 years ($p = .009$). When excluding the 17 patients from the OBS group who were operated by PTX, no difference in s-calcium levels was demonstrated from baseline to 10 years (2.61 mmol/L ± 0.10 versus 2.61 ± 0.14 mmol/L; $p = .932$). As expected, there was a significant difference in s-calcium levels between the two groups after 10 years ($p < .001$).

In the PTX group PTH declined significantly from baseline (10.16 ± 4.08 pmol/L) to 10-year values (5.44 ± 2.87 pmol/L; $p < .001$). The OBS group had stable PTH values throughout the study with baseline of 10.55 ± 4.26 pmol/L and values at 10-year visit of 10.60 ± 5.27 pmol/L. There was a significant difference in PTH values between the groups after 10 years of follow-up ($p < .001$).

SF-36

At end of study (10-year follow-up), patients randomized to PTX scored significantly better on VT ($p = .017$) compared to the OBS group (Fig. 2*A* and 2*B*). Moreover, there was a trend toward an improvement in SF following PTX ($p = .063$). Within group analyses showed that VT and SF scores improved after 10 years compared to baseline in the PTX group. All positive changes based on raw data were however modest.

In the four physical subscales, we found no treatment effect for PTX as compared to the OBS group. In general, the differences in PF, BP, GH, and RP were small and insignificant during the study (Fig. 2). Likewise, the OBS group had stable values without a significant decrease in any of the SF-36 scores throughout the study.

In the mixed model statistics, all valid SF-36 questionnaire derived data were included (Table 2). We also included data from the first-year visit (97 patients). This visit was, however, not predefined in the protocol for QoL evaluation. Therefore, we performed a reanalysis excluding these data. The results remained similar, with a significant treatment effect of PTX for VT ($p = .008$) and a borderline effect for SF ($p = .057$). In a per protocol analysis excluding the 17 patients who underwent PTX in the OBS group, there was a significant treatment effect between the groups for VT ($p = .013$) and SF ($p = .023$).

The relationship between albumin corrected s-calcium levels and SF-36 data (mixed models), showed a significant worsening of QoL scores with higher s-calcium levels in PF ($p = .021$) and RE ($p = .049$). There was no significant association between SF-36 data and PTH levels.

CPRS

After 10 years, there were no significant differences in any of the items between the two treatment groups. Both groups showed a significant decrease in the scores during the period, indicating improvement over time. For the PTX group, there was an improvement of the following symptoms: irritation ($p = .009$), pessimism ($p = .005$), unrest ($p = .031$), phobias ($p = .001$), somatization ($p = .027$), and muscle tension ($p = .025$) (Fig. 3*A*). Similarly, there was a significant improvement in the OBS group in worry ($p = .013$), unrest ($p = .010$), and somatization ($p = .032$) (Fig. 3*B*).

Compared to baseline, both groups showed a significant improvement of the CPRS sum scores at the end of study. The sum scores improved in the PTX group from a mean ± SD of 18.89 ± 1.46 to 16.09 ± 1.59 ($p = .014$). Similarly, the CPRS sum score improved in the OBS group from 19.64 ± 1.43 to 16.64 ± 1.62 ($p = .015$). The developments of sum scores for both groups are given in Fig. 3*C*. From baseline to 10 years, there was no overall treatment effect of PTX as compared to the OBS group (sum score difference: 0.198 ± 1.68) ($p = .91$) (Fig. 3*C*).

When excluding data in CPRS from year 1, the results were unchanged with no treatment effect of PTX in any of the single items or sum scores. Per protocol analysis without the 17 operated patients in OBS group had little effect on the between group differences. The sum score for the OBS group without these patients still showed an improvement from baseline to end of study ($p = .016$).

Further, we analyzed the relationship between s-calcium levels and CPRS values. There was no significant association between any of the individual questions, but a trend indicating that improvement of pessimism ($p = .065$) and concentration ($p = .082$) was related to lower s-calcium levels. When applying this model to PTH levels, we found an association between lower PTH levels and better decision making ($p = .004$). There was a trend toward an association between improved mood and appetite with lower PTH levels (mood: $p = .056$; appetite: $p = .063$).

Discussion

Based on the end of study, 10-year data, from the hitherto largest and longest-lasting RCT in mild PHPT, we show that long-term observation without intervention has no harming effect on QoL, whereas PTX leads to a modest improvement in mental subscales of SF-36, compared to observation. Moreover, we show by mixed models statistics a relationship between s-calcium levels and QoL in two SF-36 domains; ie, the higher s-calcium, the lower QoL. The CPRS data showed improvements in symptoms over the years in both groups, but without any treatment effect of PTX as compared to observation.

Patients with mild PHPT have decreased QoL compared to normal controls, even before the diagnosis has become evident.[30] Observational studies have indicated a positive effect of PTX in PHPT,[8,9] but RCT's with relatively short observation time (up to 2 years) have shown ambiguous or inconsistent results.[11–13] These findings are summarized in a recent meta-analysis,[31] where data from three randomized controlled trials (including the SIPH study) suggested that PTX leads to improvement in QoL, especially in the mental domains. Neuropsychiatric symptoms or impaired QoL were not included as indication for PTX in the latest consensus statements on management of mild PHPT.[2,7]

In the present study, the SF-36 results were remarkably stable in both groups during the whole trial period, as also seen in normative Norwegian long-term data.[32] Moreover, the within group differences over time in our study were modest, and the demonstrated beneficial effect of PTX in the mental domains were minor and of questionable clinical relevance.[32] There is no clear consensus on the minimal clinically important difference (MCID) for SF-36 data.[33,34] Generally, an improvement of 5% to 10% of the scale width is considered as an important clinical change.[35] Using the 5% cutoff, both VT (8.0%) and SF (6.5%) would be interpreted as being a clinically important improvement in this study. It is worth mentioning that none of the dimensions in any of the two groups had a clinically important negative change. A 10% cutoff is less likely to include false-positive

results. By this standard, none of the SF-36 subscales in the present study would have an MCID.

In two of the eight dimensions in SF-36, we found a relationship between s-calcium levels and QoL, where higher s-calcium values correlated to lower QoL. In the other six dimensions, no such relations could be demonstrated by the fitted linear mixed model. The relation between SF-36 scores and s-calcium levels is in accordance with findings from the validation study of the disease specific PHPQoL questionnaire,[14] and might point toward the pathogenic mechanisms of the disease. Both s-calcium and PTH levels reflect disease activity and might by themselves affect neuropsychiatric function and QoL.[22,23,36]

The CPRS scores improved modestly over the years in both groups, without any treatment effect. The modified CPRS was developed with focus on mild PHPT with focus on neuropsychiatric symptoms.[29] There is, however, still a discussion on the specificity of PHPT-related psychometric symptoms. The modified version of CPRS used for assessing neuropsychiatric symptoms in this study has numerous similarities to the later developed PHPT-specific questionnaires.[14,37] There is significant overlap between the PHPQoL and CPRS with 14 of 16 items being thematically identical. The major overlap between the modified CPRS and PHPQoL strengthens the validity of the presented data.

Because both groups improved slightly over the 10-year period without indication of a treatment effect of PTX, the change over time could be related to repetition of the questionnaire and recall bias. This does, however, not seem plausible because there were 5 years between the two last assessments. Thus, it rather reflects an age-dependent development. Although there are no long-term normative data from this modified version of CPRS, other psychometric tools have demonstrated an improvement of psychological well-being with age.[38] Whether the improvement is a linear curve from early adulthood or a U-shaped curve with improvement from about 40 to 50 years of age is debatable.[39,40] However, for our population with a mean age of 62 years at enrolment, it is likely that the improvement in the CPRS-data over a decade in both groups was due to general trends associated with aging, and not disease-specific.

We included 1-year data as an illustration of how the difference seems more pronounced at an earlier time point with a gradual convergence with time (Fig. 3*C*). The CPRS sum scores illustrate the potential placebo effect of surgical treatment, because a different course was seen in the two groups during the initial period of the study (first year). A slight improvement was observed following PTX and a more pronounced worsening by observation. Later in the study, the curves aligned and developed in parallel. It should be emphasized that these observations did not reach any statistical significance, but underscores the necessity of a control group and long-term observation when evaluating QoL in patients with mild PHPT, where short-term surgical observational studies in general favor surgical treatment.[8–10]

The strengths of the present study are the randomized design, the long-term follow-up, and the use of validated tests. Following a relatively high number of dropouts in the initial phase of the study, the adherence to the protocol was high. Nevertheless, there may have been a selection bias toward subjects with improved QoL, but the dropout rates were similar in both groups. In addition, the intention-to-treat principle for analyses is a strength, despite several patients undergoing PTX for several reasons during the study. We used mixed model statistics enabling all data points to be included, thereby increasing the sensitivity of the statistical methods. The limitations are, however, also numerous. First, the study was open, although randomized. Several patients in the initial phase of the study did not accept the randomization and withdrew. In addition, when analyzing QoL data, several variables were analyzed and multiplicity could influence the statistical results for single variables. CRPS has 17 single items (and two sum scores) and correction for multiple comparisons could be argued. Because our overall conclusion is that there are no differences between the groups, we have presented data without correction for multiple comparisons. Because the primary endpoint of the SIPH study was mortality, no power calculations for QoL scores were performed and thus, a type two error cannot be excluded. It should be emphasized that the presented data do not include an evaluation of cognitive function, another important aspect of PHPT.[41] On the other hand, the present study is unique evaluating the effect of PTX in a randomized design over one decade.

In conclusion, in the span of 10 years following randomization to observation without intervention versus PTX in mild PHPT, QoL data by SF-36 and neuropsychiatric scores by the modified CPRS, were remarkably stable in both groups. Only subtle benefits of surgery were found in the mental subscales in SF-36, of uncertain clinical significance. Thus, from a QoL perspective it seems safe to observe patients with mild PHPT for several years.

Disclosures

All authors declare that they have no potential conflicts of interest.

Acknowledgments

For this work, the SIPH-study received grants from the South-Eastern Norway Regional Health Authority and the Swedish Government grants under the ALF-agreement (ALF VGR-71160).

Authors' roles: Conceptualization and funding acquisition: JB. Investigation: KG, CM, SLF, YP, TA, OH, TR, JN, SJ, AH, and JB. Methodology and data analysis: MP, MWF, and JB. Project administration and supervision: KG, MH, AH, and JB. Validation: MP, KG, and AH. Visualization: MP. Writing original draft: MP, KL, MWF, AH, and JB. All authors have contributed with reviewing and editing of original draft and have approved the manuscript for publication.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1002/jbmr.4199.

References

1. Bollerslev J, Schalin-Jäntti C, Rejnmark L, et al. Management of endocrine disease: unmet therapeutic, educational and scientific needs in parathyroid disorders. Eur J Endocrinol. 2019;181(3):P1–19.

2. Khan AA, Hanley DA, Rizzoli R, et al. Primary hyperparathyroidism: review and recommendations on evaluation, diagnosis, and management. A Canadian and international consensus. Osteoporos Int. 2017;28(1):1–19.
3. Bandeira F, Griz L, Caldas G, Bandeira C, Freese E. From mild to severe primary hyperparathyroidism: the Brazilian experience. Arq Bras Endocrinol Metabol. 2006;50(4):657–63.
4. Bilezikian JP, Meng X, Shi Y, Silverberg SJ. Primary hyperparathyroidism in women: a tale of two cities—New York and Beijing. Int J Fertil Womens Med. 2000;45(2):158–65.
5. Shah VN, Bhadada S, Bhansali A, Behera A, Mittal BR. Changes in clinical & biochemical presentations of primary hyperparathyroidism in India over a period of 20 years. Indian J Med Res. 2014;139(5):694–9.
6. Zhao L, Liu JM, He XY, et al. The changing clinical patterns of primary hyperparathyroidism in Chinese patients: data from 2000 to 2010 in a single clinical center. J Clin Endocrinol Metab. 2013;98(2):721–8.
7. Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the Fourth International Workshop. J Clin Endocrinol Metab. 2014;99(10):3561–9.
8. Bannani S, Christou N, Guerin C, et al. Effect of parathyroidectomy on quality of life and non-specific symptoms in normocalcaemic primary hyperparathyroidism. Br J Surg. 2018;105(3):223–9.
9. Weber T, Eberle J, Messelhauser U, et al. Parathyroidectomy, elevated depression scores, and suicidal ideation in patients with primary hyperparathyroidism: results of a prospective multicenter study. JAMA Surg. 2013;148(2):109–15.
10. Storvall S, Ryhänen EM, Heiskanen I, Sintonen H, Roine RP, Schalin-Jäntti C. Surgery significantly improves neurocognition, sleep, and blood pressure in primary hyperparathyroidism: a 3-year prospective follow-up study. Hormone Metab Res. 2017;49(10):772–7.
11. Ambrogini E, Cetani F, Cianferotti L, et al. Surgery or surveillance for mild asymptomatic primary hyperparathyroidism: a prospective, randomized clinical trial. J Clin Endocrinol Metab. 2007;92(8):3114–21.
12. Bollerslev J, Jansson S, Mollerup CL, et al. Medical observation, compared with parathyroidectomy, for asymptomatic primary hyperparathyroidism: a prospective, randomized trial. J Clin Endocrinol Metab. 2007;92(5):1687–92.
13. Rao DS, Phillips ER, Divine GW, Talpos GB. Randomized controlled clinical trial of surgery versus no surgery in patients with mild asymptomatic primary hyperparathyroidism. J Clin Endocrinol Metab. 2004; 89(11):5415–22.
14. Webb SM, Puig-Domingo M, Villabona C, et al. Validation of PHPQoL, a disease-specific quality-of-life questionnaire for patients with primary hyperparathyroidism. J Clin Endocrinol Metab. 2016;101(4):1571–8.
15. Ejlsmark-Svensson H, Sikjaer T, Webb SM, Rejnmark L, Rolighed L. Health-related quality of life improves 1 year after parathyroidectomy in primary hyperparathyroidism: a prospective cohort study. Clin Endocrinol. 2019;90(1):184–91.
16. Elvius M, Lagrelius A, Nygren A, Alveryd A, Christensson TA, Nordenstrom J. Seventeen year follow-up study of bone mass in patients with mild asymptomatic hyperparathyroidism some of whom were operated on. Eur J Surg. 1995;161(12):863–9.
17. Rubin MR, Bilezikian JP, McMahon DJ, et al. The natural history of primary hyperparathyroidism with or without parathyroid surgery after 15 years. J Clin Endocrinol Metab. 2008;93(9):3462–70.
18. Lundstam K, Heck A, Godang K, et al. Effect of surgery versus observation: Skeletal 5-year outcomes in a randomized trial of patients with primary HPT (the SIPH Study). J Bone Miner Res. 2017;32(9): 1907–14.
19. Parfitt AM. Large calcium fluxes that are not related to remodeling exist. Bone. 2003;33(3):269.
20. Lupp A, Klenk C, Röcken C, Evert M, Mawrin C, Schulz S. Immunohistochemical identification of the PTHR1 parathyroid hormone receptor in normal and neoplastic human tissues. Eur J Endocrinol. 2010; 162(5):979–86.
21. Usdin TB, Gruber C, Bonner TI. Identification and functional expression of a receptor selectively recognizing parathyroid hormone, the PTH2 receptor. J Biol Chem. 1995;270(26):15455–8.
22. Bago AG, Dimitrov E, Saunders R, et al. Parathyroid hormone 2 receptor and its endogenous ligand tuberoinfundibular peptide of 39 residues are concentrated in endocrine, viscerosensory and auditory brain regions in macaque and human. Neuroscience. 2009;162(1): 128–47.
23. Bollerslev J, Sjöstedt E, Rejnmark L. Cardiovascular consequences of parathyroid disorders in adults. Ann Endocrinol. 2020;4266(20): 30026–3.
24. Potts JTAI, Barker CF, Brennan MF, Coburn JW, Wong GL. Diagnosis and management of asymptomatic primary hyperparathyroidism: Consensus Development Conference Statement. Ann Internal Med. 1991;114(7):593–7.
25. Ware JE Jr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) project. J Clin Epidemiol. 1998;51(11):903–12.
26. Sullivan M, Karlsson J, Ware JE Jr. The Swedish SF-36 Health Survey—I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. Soc Sci Med. 1995;41(10):1349–58.
27. Asberg M, Montgomery SA, Perris C, Schalling D, Sedvall G. A comprehensive psychopathological rating scale. Acta Psychiatr Scand Suppl. 1978;271:5–27.
28. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382–9.
29. Lundgren E, Ljunghall S, Akerstrom G, Hetta J, Mallmin H, Rastad J. Case-control study on symptoms and signs of "asymptomatic" primary hyperparathyroidism. Surgery. 1998;124(6):980–5.
30. Siilin H, Rastad J, Ljunggren O, Lundgren E. Disturbances of calcium homeostasis consistent with mild primary hyperparathyroidism in premenopausal women and associated morbidity. J Clin Endocrinol Metab. 2008;93(1):47–53.
31. Cheng SP, Lee JJ, Liu TP, et al. Quality of life after surgery or surveillance for asymptomatic primary hyperparathyroidism: a meta-analysis of randomized controlled trials. Medicine. 2015;94(23): e931.
32. Jacobsen EL, Bye A, Aass N, et al. Norwegian reference values for the Short-Form Health Survey 36: development over time. Qual Life Res. 2018;27(5):1201–12.
33. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407–15.
34. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. 1989;27(3 Suppl): S178–89.
35. Osoba D, Bezjak A, Brundage M, Zee B, Tu D, Pater J. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of the National Cancer Institute of Canada Clinical Trials Group. Eur J Cancer. 2005;41(2):280–7.
36. Walker MD, Silverberg SJ. Primary hyperparathyroidism. Nat Rev Endocrinol. 2018;14(2):115–25.
37. Pasieka JL, Parsons LL. Prospective surgical outcome study of relief of symptoms following surgery in patients with primary hyperparathyroidism. World J Surg. 1998;22(6):513–8 discussion 8-9.
38. Nilsson KW, Leppert J, Simonsson B, Starrin B. Sense of coherence and psychological well-being: improvement with age. J Epidemiol Community Health. 2010;64(4):347–52.
39. Blanchflower DG, Oswald AJ. Is well-being U-shaped over the life cycle? Soc Sci Med. 2008;66(8):1733–49.
40. Thomas ML, Kaufmann CN, Palmer BW, et al. Paradoxical trend for improvement in mental health with aging: a community-based study of 1,546 adults aged 21-100 years. J Clin Psychiatry. 2016;77(8):e1019–25.
41. Walker MD, McMahon DJ, Inabnet WB, et al. Neuropsychological features in primary hyperparathyroidism: a prospective study. J Clin Endocrinol Metab. 2009;94(6):1951–8.

II

III

CLINICAL TRIAL

JBMR®

Positive Effect of Parathyroidectomy Compared to Observation on BMD in a Randomized Controlled Trial of Mild Primary Hyperparathyroidism

Karolina Lundstam,[1,2] Mikkel Pretorius,[3,4] Jens Bollerslev,[3,4] Kristin Godang,[3] Morten W Fagerland,[5] Charlotte Mollerup,[6] Stine L Fougner,[7] Ylva Pernow,[8] Turid Aas,[9] Ola Hessman,[10] Thord Rosén,[11] Jörgen Nordenström,[12] Svante Jansson,[13] Mikael Hellström,[1,2] and Ansgar Heck[3,4]

[1]Department of Radiology, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[2]Department of Radiology, Sahlgrenska University Hospital, Gothenburg, Sweden
[3]Section of Specialized Endocrinology, Oslo University Hospital, Oslo, Norway
[4]Faculty of Medicine, University of Oslo, Oslo, Norway
[5]Oslo Centre for Biostatistics and Epidemiology, Research Support Services, Oslo University Hospital, Oslo, Norway
[6]Clinic of Breast and Endocrine Surgery, Center HOC, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
[7]Department of Endocrinology, Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
[8]Department of Molecular Medicine and Surgery, Department of Endocrinology, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
[9]Department of Breast and Endocrine Surgery, Haukeland University Hospital, Bergen, Norway
[10]Department of Surgery and Centre for Clinical Research of Uppsala University, Västmanland's Hospital, Västerås, Sweden
[11]Department of Medicine, Section of Endocrinology, Diabetes, and Metabolism, Sahlgrenska University Hospital, Gothenburg, Sweden
[12]Department of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
[13]Department of Surgery, Institute of Clinical Sciences, Sahlgrenska Academy at the University of Gothenburg, Sahlgrenska University Hospital, Gothenburg, Sweden

ABSTRACT

Mild or asymptomatic disease is now the dominating presentation of primary hyperparathyroidism (PHPT). However, bone involvement with decreased bone mineral density (BMD) and an increased risk of fractures has been demonstrated. Indications for parathyroidectomy (PTX) in mild PHPT have been debated for years. There is a need of long-term randomized studies comparing PTX with observation without intervention (OBS). Here, we present bone health data from the Scandinavian Investigation of Primary Hyperparathyroidism (SIPH), a randomized controlled trial, comparing PTX to OBS. This study included 191 patients (96 OBS/95 PTX), and 129 patients (64 OBS/65 PTX) were followed for 10 years to the end of study (EOS). BMD was measured with dual-energy X-ray absorptiometry (DXA), peripheral fractures were noted, and spine radiographs were obtained for vertebral fracture assessment. There was a significant treatment effect of PTX on BMD compared with OBS for all analyzed compartments, most explicit for the lumbar spine (LS) and femoral neck (FN) ($p < 0.001$). The mean changes in *T*-score from baseline to 10 years were from 0.41 for radius 33% (Rad33) to 0.58 for LS greater in the PTX group than in the OBS group. There was a significant decrease in BMD for all compartments in the OBS group, most pronounced for FN, Rad33, and ultradistal radius (UDR) ($p < 0.001$). Even though there was a significant treatment effect of PTX compared with OBS, there was only a significant increase in BMD over time for LS ($p < 0.001$). We found no difference between groups in fracture frequency in the 10-year cohort, neither with modified intention-to-treat (mITT) analysis nor per protocol analysis. Because BMD is only a surrogate endpoint of bone health and PTX did not reduce fracture risk, observation could be considered a safe option for many patients with mild PHPT regarding bone health in a 10-year perspective. © 2023 The Authors. *Journal of Bone and Mineral Research* published by Wiley Periodicals LLC on behalf of American Society for Bone and Mineral Research (ASBMR).

Received in original form June 29, 2022; revised form November 25, 2022; accepted December 29, 2022.

Address correspondence to: Karolina Lundstam, MD, Department of Radiology, Sahlgrenska University Hospital, S-413 45, Gothenburg, Sweden. E-mail: karolina.lundstam@vgregion.se

Additional Supporting Information may be found in the online version of this article.

Karolina Lundstam and Mikkel Pretorius share first authorship.

Mikael Hellström and Ansgar Heck share last authorship.

The original study protocol from 1998 (translated to English) is included as a supplement (Supplemental File S1). Access to de-identified data from the SIPH trial will be made available upon e-mail request to the corresponding author. Data will be shared within the Scandinavian laws and legislations.

Journal of Bone and Mineral Research, Vol. 00, No. 00, Month 2023, pp 1–9.

DOI: 10.1002/jbmr.4763

KEY WORDS: PARATHYROID-RELATED DISORDERS; CLINICAL TRIALS; DXA; FRACTURE PREVENTION; RADIOLOGY

Introduction

Primary hyperparathyroidism (PHPT) is potentially a serious disease with multi-organ involvement. However, in developed countries, PHPT is now most common in patients with only mild or no obvious symptoms, due to more frequent detection by routine clinical testing of calcium levels as part of multichannel blood testing. The only curative treatment is parathyroidectomy (PTX). Although there is consensus that patients with symptomatic PHPT should be offered PTX,[1] there are uncertainties regarding the optimal management of patients with mild PHPT.[2] This is reflected in four subsequent international workshops addressing criteria for surgery in mild PHPT, with extended indications over time.[3-6]

PHPT is known to affect bone health. A continuous excess of parathyroid hormone (PTH) increases bone turnover, potentially leading to a decreased bone mineral density (BMD).[7,8] Conversely, several studies have shown increased BMD after PTX.[8-11] A long-term observational study found the BMD increase after PTX to persist for up to 15 years. The same study also described stable lumbar spine BMD in patients without PTX but a significant decrease in femoral neck and radial BMD starting just before 10 years of follow-up.[12] Two randomized studies comparing PTX with observation (OBS), with a follow-up time of 1 to 2 years, have demonstrated positive effects of PTX on lumbar spine and femoral neck BMD but no treatment effect on radial BMD compared with OBS.[13,14]

There is, however, a lack of long-term randomized studies, which are needed for deeper knowledge on how PTX affects patients with mild PHPT over time.

We have previously presented 5-year BMD data from our prospective, randomized, controlled study on patients with mild PHPT, Scandinavian Investigation of Primary Hyperparathyroidism (SIPH), comparing PTX with OBS.[15] A treatment effect of PTX, compared with OBS, on bone mineral density was found in four of five DXA compartments,[15] but there was no significant difference in the frequency of vertebral and peripheral fractures between the groups.[16]

The ultimate aim of the randomized SIPH study was to compare the long-term (10-year) effects of PTX compared with OBS. Recently, fracture data were presented, showing no difference in fracture frequency between the randomization groups.[17] Here, we present the 10-year BMD data.

Materials and Methods

Participants

Between 1998 and 2005, 191 patients meeting the NIH 1991 criteria for mild PHPT[3] were included in the SIPH study in Sweden, Norway, and Denmark (ClinicalTrials.gov: NCT00522028). They were randomized (in blocks of 2 to 10 patients at each center) either to PTX (95 patients; 13 men) or OBS (96 patients; 13 men). Inclusion criteria were untreated mild PHPT, no thiazide medication within 4 weeks of calcium and PTH measurements for diagnosis, age 50 to 80 years, albumin-corrected calcium between 2.60 and 2.80 mmol/L, and informed consent. Exclusion criteria were previous neck surgery, osteitis fibrosa cystica, serum-creatinine >130 μmol/L, kidney stones, complicated medical conditions, psychiatric disorders, and familial forms of hyperparathyroidism. The study has previously been described in detail.[18]

According to the protocol, the end-of-study (EOS) visit was planned to occur 10 years after inclusion. Of the 191 patients included, 145 patients remained in the study at the 5-year visit and 129 patients concluded the 10-year visit. In the 10-year cohort, 65 patients (9 men) had been randomized to PTX and 64 (7 men) to OBS. No long-term surgical complications (chronic hypocalcemia or recurrent laryngeal nerve palsy) of PTX were revealed. Reasons for early termination and withdrawals, as well as the number of patients included in the EOS analyses, are shown in Fig. 1.

Seventeen patients (17.7%) in the OBS group underwent PTX within the 10-year study period, 4 due to medical predefined protocol criteria and the rest due to patient's decision or non-protocol specified medical conditions. Thirteen of these concluded the EOS visit (20.3% of the 129 patients concluding the EOS visit, ie, of the 10-year cohort). According to a modified intention-to-treat (mITT; including only the 129 patients in the 10-year cohort) principle, these patients remained in the OBS group for statistical analyses.

The study was approved by the local ethics committees in the respective countries and was conducted according to the Declaration of Helsinki II. All participants gave informed consent.

Biochemistry

Calcium, albumin, and PTH blood levels were measured locally at each center at baseline and at the yearly visits. Data management and albumin correction of calcium levels were performed at the coordinating study center (Oslo University Hospital), and the reported biochemical measurements from baseline visits, 5-year visits, and 10-year visits in the case report forms (CRFs) have subsequently been validated to find and correct for reporting errors.

Dual-energy X-ray absorptiometry (DXA)

BMD was measured with DXA at baseline and after 5 and 10 years. Patients were scanned at several different centers over a time span of almost 19 years. Standardized calibrating procedures were applied by all centers. Printouts of all scans were sent to the coordinating center. The dominating manufacturer was Lunar (GE Healthcare, Madison, WI, USA) (270 scans, 106 patients), with six different hardware used (Expert, DPX-L, DPX-IQ, Prodigy, Prodigy Advance, and iDXA). Scanners from Norland (Norland Corp., Fort Atkinson, WI, USA) (27 scans, 9 patients) and Hologic (Hologic Inc., Waltham, MA, USA) (31 scans, 12 patients) were used at one and two centers, respectively.

The BMD–*T*-score plot by Lundstam and colleagues, a simple tool for detecting random (eg, typos) and systematic BMD data irregularities, based on analysis of the BMD–*T*-score relationship,[19] was used for DXA BMD data validation from lumbar spine (LS), femoral neck (FN), radius 33% (Rad33), and ultradistal radius (UDR). A description of this process is found in Supplemental File S2.

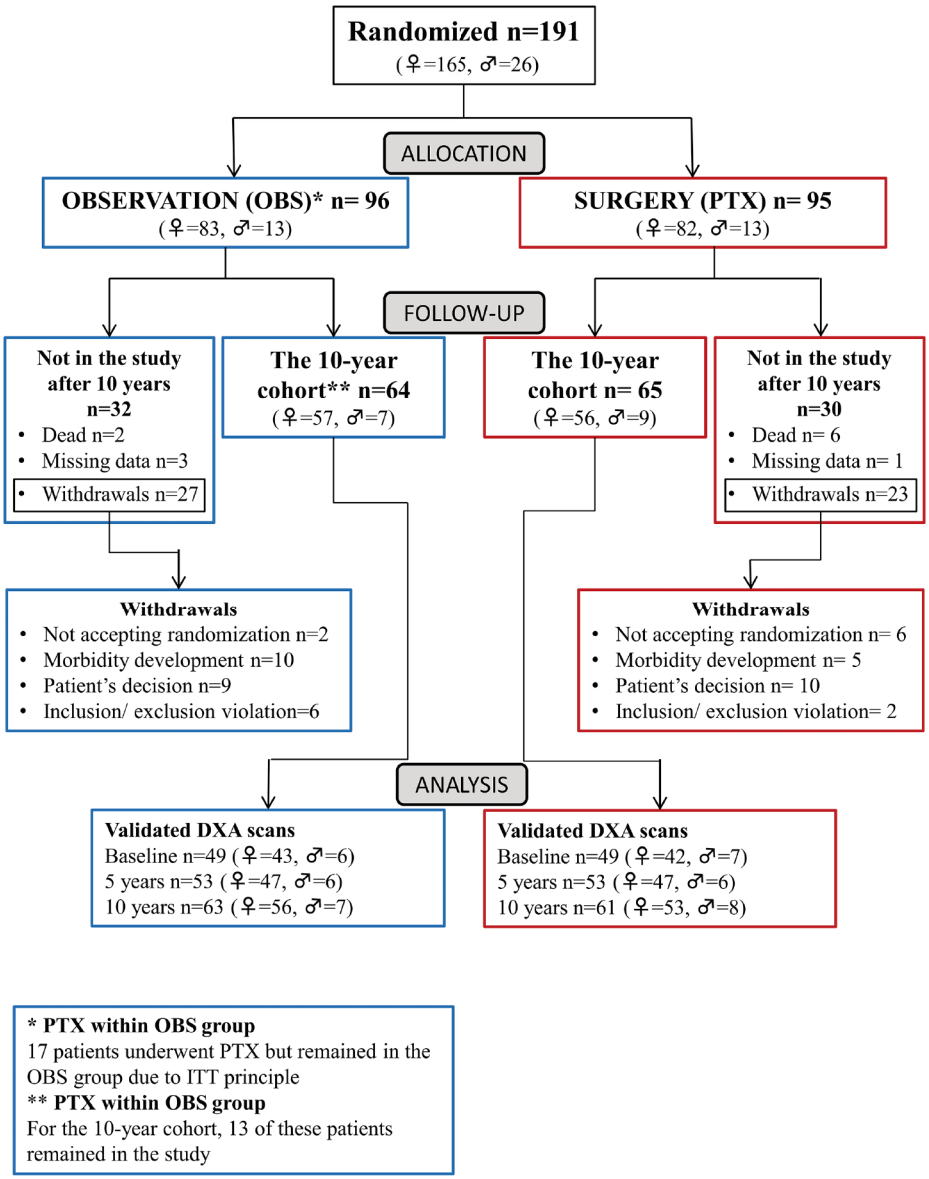

Fig. 1. CONSORT (consolidated standards of reporting trials) flow diagram of the patient population in the Scandinavian Investigation of Primary Hyperparathyroidism (SIPH) study. ♀ = females; ♂ = males; DXA = dual-energy X-ray absorptiometry; ITT = intention to treat.

ITT and per protocol analysis of any fracture in the 10-year cohort

Recently, new (incidental) vertebral and peripheral fractures, occurring after study start, were reported together with the primary endpoint of the SIPH study, mortality.[17] Here, we additionally present an overall analysis of patients with any new fracture in the 10-year cohort with both mITT analysis and per protocol analysis.

Statistical analysis

Depending on data distribution, mean ± standard deviation (SD) or median (interquartile range [IQR]) were used. Differences within and between groups regarding change in calcium and PTH levels (baseline to EOS), were analyzed with dependent and independent samples *t* test, respectively, as analyzed data were normally distributed. A linear mixed model was fitted to the DXA measurements from the different time points (baseline, 5 years, and 10 years). Based on the model fit, mean values for the randomization groups at the different time points, mean values for within-group change between baseline and 10 years, and mean values for treatment effect of PTX compared with OBS, from baseline to 10 years, were estimated with 95% confidence interval (95% CI). Chi-square test was used for comparing randomization groups with regard to development of any new fracture or not. Stata/SE 16.1 (StataCorp LLC, College Station,

Table 1. Baseline Characteristics of the 10-Year Cohort ($n = 129$)

	OBS ($N = 64$)	PTX ($N = 65$)
Sex		
Female, *n* (%)	57 (89.1)	56 (86.2)
Male, *n* (%)	7 (10.9)	9 (13.8)
Age (years) at inclusion, mean ± SD	61.5 ± 6.5	60.6 ± 6.9
Calcium[a] (mmol/L), mean ± SD	2.63 ± 0.11	2.65 ± 0.12
PTH (pmol/L), median (IQR)	9.3 (4.6)	9.5 (5.1)
Lumbar spine, mean ± SD		
BMD (g/cm^2) (OBS = 48; PTX = 47)	1.060 ± 0.19	1.081 ± 0.18
T-score (OBS = 48; PTX = 48)	−1.24 ± 1.63	−1.08 ± 1.49
Femoral neck, mean ± SD		
BMD (g/cm^2) (OBS = 46; PTX = 43)	0.835 ± 0.11	0.834 ± 0.13
T-score (OBS = 46; PTX = 44)	−1.29 ± 0.89	−1.36 ± 1.12
Radius 33%, mean ± SD		
BMD (g/cm^2) (OBS = 38; PTX = 35)	0.618 ± 0.11	0.647 ± 0.09
T-score (OBS = 38; PTX = 35)	−1.47 ± 1.36	−1.08 ± 1.17
Ultradistal radius, mean ± SD		
BMD (g/cm^2) (OBS = 38; PTX = 34)	0.310 ± 0.08	0.308 ± 0.07
T-score (OBS = 38; PTX = 34)	−1.98 ± 1.95	−2.05 ± 1.63

Abbreviations: OBS = observation; PTX = parathyroidectomy; SD = standard deviation; IQR = interquartile range; BMD = bone mineral density.
[a]Albumin-corrected calcium.

TX, USA) was used for the statistical analysis. A p value <0.05 was considered statistically significant.

Results

Baseline characteristics and biochemistry

Data presented here are from the 10-year cohort, the 129 patients who completed the EOS visit. This is in concordance with our previously published results from the patients concluding the 5-year visit.[15,16] Baseline characteristics for the 10-year cohort are shown in Table 1.

There were treatment effects on both albumin-corrected calcium and PTH levels in the 10-year perspective ($p < 0.001$ and $p < 0.001$, respectively). Even so, albumin-corrected calcium decreased over time in both groups (OBS: mean difference [diff] −0.09 ± 0.23 mmol/L, $p = 0.007$; PTX: mean diff −0.30 ± 0.20 mmol/L, $p < 0.001$). However, if the patients in the OBS group who underwent PTX were excluded, there was no significant change in calcium levels in the OBS group (Fig. 2). There was a decrease in PTH levels over 10 years in the PTX group (mean diff −4.64 ± 5.04 pmol/L, $p < 0.001$) but no significant change in the OBS group.

DXA

There were treatment effects of PTX compared with OBS for all analyzed compartments with mixed model statistics of BMD and *T*-scores from baseline to EOS (FN and LS: $p < 0.001$; forearm: $p = 0.006$ to $p = 0.012$), as shown in Table 2. At EOS, the between-group differences in interval change in *T*-scores (PTX group compared with the OBS group) ranged from 0.41 (Rad33) to 0.58 (LS). *T*-score differences over time for the two groups are illustrated in Fig. 3.

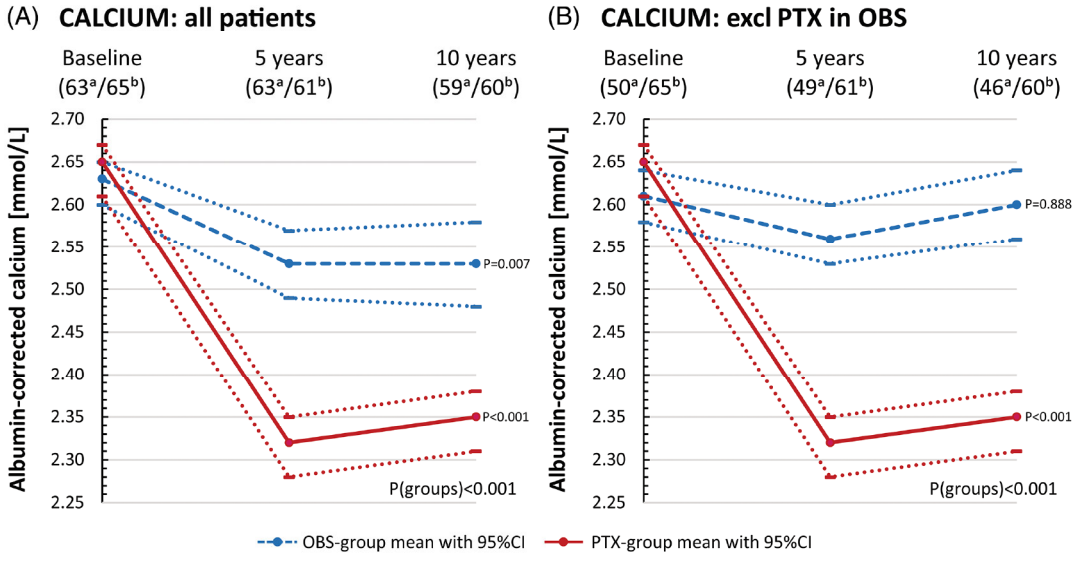

Fig. 2. Albumin-corrected calcium over time for (*A*) all 129 patients in the 10-year cohort and (*B*) the same after exclusion of those in the OBS group that underwent PTX. P denotes p value for longitudinal changes from baseline to 10 years within each group. P(groups) denotes p value for longitudinal changes between groups from baseline to 10 years. OBS = observation; PTX = parathyroidectomy. [a]Number of measurements of albumin-corrected calcium in the OBS group. [b]Number of measurements of albumin-corrected calcium in the PTX group.

Table 2. DXA Results From Baseline to 10 Years With Treatment Effects of Parathyroidectomy Derived From Mixed Model Statistics

DXA compartment	Baseline Mean (95% CI)	5 years Mean (95% CI)	10 years Mean (95% CI)	Change from baseline to 10 years Mean (95% CI)	Interval change from baseline to 10 years, between-group difference Mean (95% CI); *p* value
LS BMD					
OBS ($n = 63$)	1.066 (1.018 to 1.113)	1.054 (1.009 to 1.100)	1.043 (0.997 to 1.090)	−0.022 (−0.044 to −0.001)**	0.069 (0.038 to 0.100); $p < 0.001$
PTX ($n = 62$)	1.090 (1.042 to 1.137)	1.113 (1.067 to 1.159)	1.136 (1.089 to 1.183)	0.046 (0.024 to 0.068)*	
LS *T*-score					
OBS ($n = 63$)	−1.19 (−1.58 to −0.80)	−1.30 (−1.67 to −0.93)	−1.41 (−1.79 to −1.03)	−0.22 (−0.40 to −0.04)**	0.58 (0.32 to 0.83); $p < 0.001$
PTX ($n = 62$)	−1.00 (−1.39 to −0.61)	−0.82 (−1.20 to −0.44)	−0.64 (−1.02 to −0.26)	0.36 (0.18 to 0.54)*	
FN BMD					
OBS ($n = 63$)	0.840 (0.811 to 0.868)	0.807 (0.780 to 0.834)	0.775 (0.747 to 0.803)	−0.065 (−0.080 to −0.049)*	0.050 (0.028 to 0.072); $p < 0.001$
PTX ($n = 60$)	0.849 (0.819 to 0.878)	0.841 (0.813 to 0.869)	0.834 (0.805 to 0.862)	−0.015 (−0.031 to 0.001)	
FN *T*-score					
OBS ($n = 63$)	−1.26 (−1.50 to −1.03)	−1.54 (−1.76 to −1.31)	−1.81 (−2.03 to −1.58)	−0.54 (−0.67 to −0.41)*	0.44 (0.26 to 0.63); $p < 0.001$
PTX ($n = 60$)	−1.23 (−1.47 to −0.99)	−1.28 (−1.50 to −1.05)	−1.33 (−1.56 to −1.10)	−0.10 (−0.23 to 0.04)	
Rad33 BMD					
OBS ($n = 50$)	0.608 (0.580 to 0.637)	0.577 (0.549 to 0.604)	0.545 (0.517 to 0.573)	−0.063 (−0.078 to −0.049)*	0.029 (0.008 to 0.050); $p = 0.008$
PTX ($n = 49$)	0.644 (0.615 to 0.673)	0.626 (0.599 to 0.654)	0.609 (0.581 to 0.637)	−0.035 (−0.050 to −0.019)*	
Rad33 *T*-score					
OBS ($n = 50$)	−1.58 (−1.94 to −1.22)	−2.02 (−2.36 to −1.67)	−2.45 (−2.81 to −2.10)	−0.87 (−1.07 to −0.67)*	0.41 (0.12 to 0.70); $p = 0.006$
PTX ($n = 49$)	−1.10 (−1.46 to −0.73)	−1.33 (−1.67 to −0.98)	−1.56 (−1.91 to −1.20)	−0.46 (−0.67 to −0.25)*	
UDR BMD					
OBS ($n = 50$)	0.305 (0.286 to 0.325)	0.289 (0.271 to 0.308)	0.273 (0.254 to 0.292)	−0.032 (−0.043 to −0.021)*	0.020 (0.004 to 0.036); $p = 0.012$
PTX ($n = 49$)	0.314 (0.294 to 0.334)	0.308 (0.289 to 0.327)	0.302 (0.283 to 0.322)	−0.012 (−0.023 to −0.0004)**	
UDR *T*-score					
OBS ($n = 50$)	−2.08 (−2.57 to −1.58)	−2.50 (−2.97 to −2.04)	−2.92 (−3.41 to −2.44)	−0.85 (−1.14 to −0.56)*	0.54 (0.12 to 0.96); $p = 0.012$
PTX ($n = 49$)	−1.84 (−2.35 to −1.34)	−2.00 (−2.47 to −1.53)	−2.15 (−2.64 to −1.66)	−0.31 (−0.61 to −0.002)**	

Abbreviations: DXA = dual-energy X-ray absorptiometry; CI = confidence interval; LS = lumbar spine; FN = femoral neck, Rad33 = radius 33%; UDR = ultradistal radius; BMD = bone mineral density (g/cm^2); OBS = observation; PTX = parathyroidectomy.

*$p < 0.001$.

**$p < 0.05$ (and >0.01).

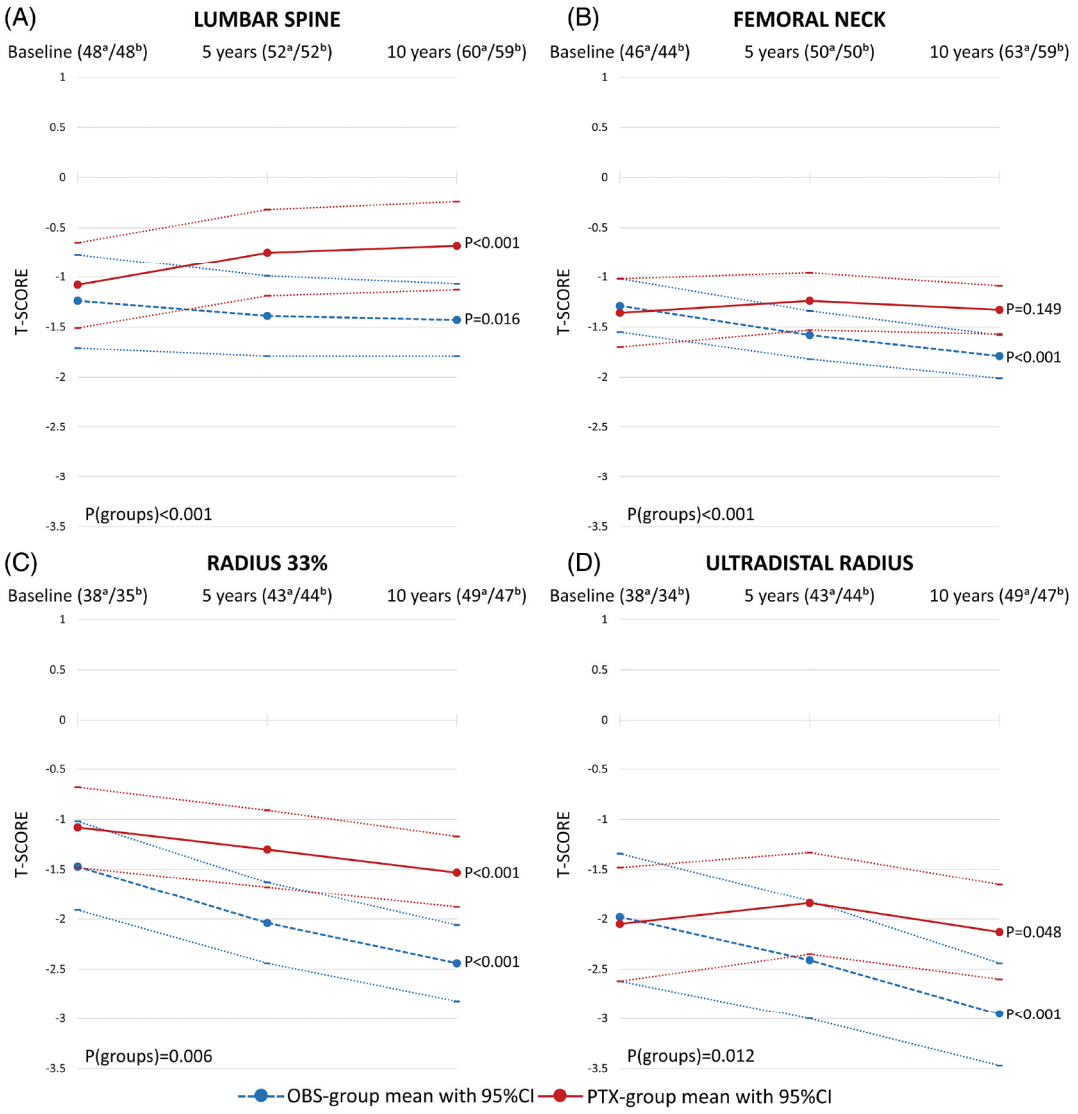

Fig. 3. Mean *T*-scores with 95% confidence intervals (95% CI) calculated from raw data (not mixed model derived data) for both randomization groups at baseline, 5 years, and 10 years for (*A*) lumbar spine, (*B*) femoral neck, (*C*) radius 33%, and (*D*) ultradistal radius. P denotes *p* value for longitudinal changes from baseline to 10 years within each group. P(groups) denotes *p* value for longitudinal changes between groups from baseline to 10 years. OBS = observation; PTX = parathyroidectomy. [a]Number of validated scans for each compartment in the OBS group. [b]Number of validated scans for each compartment in the PTX group.

A decrease in BMD and *T*-score was noted for all compartments in the OBS group (Table 2, Fig. 3). The mean *T*-score change ranged from −0.87 (Rad33: $p < 0.001$) to −0.22 (LS: $p = 0.045$) (Table 2). BMD and *T*-scores also decreased in the forearm compartments in the PTX group (Rad33 BMD: $p < 0.001$; Rad33 *T*-score: $p < 0.001$; UDR BMD: $p = 0.043$; UDR *T*-score: $p = 0.048$). The only compartment with a *T*-score increase in the PTX group was LS (0.36; $p < 0.001$; Table 2).

New (incidental) fractures in the 10-year cohort, as per ITT and per protocol

After exclusion of one patient with an indeterminate vertebral fracture,[17] there were, in the 10-year cohort, 21 of 64 patients (32.8%) in the OBS group and 18 of 64 patients (28.1%) in the PTX group with any (peripheral or vertebral) new fracture, with no difference between groups ($p = 0.565$; mITT). If the patients with PTX in the OBS group were excluded, 17 of 51 patients (33.3%) had any new fracture, with no difference between groups ($p = 0.546$; per protocol).

Discussion

In this randomized long-term study on mild PHPT comparing PTX with OBS, we found a significant PTX treatment effect on BMD for all analyzed DXA compartments. However, this treatment effect did not lead to a lower incidence of vertebral or peripheral fractures in patients randomized to PTX, as fracture events were similar between randomization groups.[17] Because fracture outcome is a hard endpoint with relevance for the patient, observation of patients with mild PHPT can be considered safe in a 10-year perspective, with no proven practical bone health benefits of PTX. However, decreasing BMD may be a concern in a longer perspective.

In accordance with our previously published 5-year BMD data,[15] a significant PTX treatment effect compared with OBS was found for LS, FN, and UDR. Cortical bone is known for poor recovery after PTX,[20,21] and our results, showing a decrease in Rad33 BMD in the PTX group, support this. However, the BMD decrease curve appears flatter for the PTX group than for the OBS group after 10 years, resulting in a significant difference between the groups, in contrast to our 5-year data where no difference was found.[15] It seems that the positive effect of PTX compared with OBS on BMD increases with time. Other randomized studies with shorter follow-up times found a positive BMD treatment effect of PTX only on LS and the femur compartments.[13,14] Our previously published 5-year data from the SIPH study found a positive treatment effect of PTX compared with OBS on LS, FN, UDR, and TB BMD.[15] The 10-year data reported in this article could add Rad33 to the list of compartments with a PTX treatment effect compared with OBS. The graphical illustration of *T*-scores over time (Fig. 3) supports this conclusion. We found a significant decrease in BMD in the OBS group for all compartments in a 10-year perspective. The decrease was most pronounced for the forearm compartments with a mean decrease of almost one *T*-score unit. There are only a few observational studies with follow-up for 10 years, all with very limited numbers of patients.[12,20,22] Thus, the idea of BMD being stable in observed patients with PHPT are based on either a short follow-up time or very few followed patients, and our results contradict this idea. There are difficulties when comparing studies, as patients with PHPT do not constitute a homogenous group. For example, included patients may be symptomatic and/or asymptomatic, an obvious cause of inhomogeneity. In addition, PHPT presentation differs between postmenopausal women compared with premenopausal women and men,[23] and there are sex differences in patients diagnosed after the age of 50 years, with more osteoporosis in females.[24]

The present study was based on the NIH Consensus meeting from 1991,[3] but since then, criteria for surgery have been extended. Based on the *T*-score ≤ −2.5 at LS, FN, or Rad33 or presence of a vertebral fracture at baseline, 38 (20 OBS/18 PTX) of our patients in the 10-year cohort (29.5%) would have met the 2014 criteria for surgery.[6] The majority met the updated criteria due to low BMD. However, no difference in fracture outcome was found among those meeting the updated surgical criteria; eight new fractures in the OBS group and seven in the PTX group, $p = 0.942$ (one patient with indeterminate vertebral fracture excluded). Interestingly, the positive BMD effect of PTX did not translate into reduced fracture rates compared with OBS, which could have been expected. Per protocol analysis of fracture frequency, with exclusion of patients in the OBS group who underwent PTX, did not change that. Possible causes include type 2 error or BMD representing bone density and not bone strength. Previous studies have found BMD to be less predictive of fractures in PHPT patients,[25,26] supporting the latter. A recent large-scale, retrospective observational study found PTX to be associated with reduced fracture risk, supporting the former.[27] However, as rightly pointed out by the authors, the retrospective design by itself, the lack of biochemical and bone mineral density data, and uncertainty about the severity of PHPT in the included patients make their conclusions uncertain. As the BMD effect of PTX compared with OBS seems to increase with time, a future difference in fracture outcome might appear. There was only one hip fracture in our study population,[17] a classical fragility fracture.[28,29] In a recent Gothenburg register-based study, hip fractures were the second most common fracture, with a mean patient age at time of fracture of 81.1 years.[29] The low number of hip fractures in our study is likely explained by our study population being younger. Therefore, one can assume that more patients in our study population will suffer from hip fractures in the future. The overall fracture incidence increases with age,[29] and with the BMD difference between PTX and OBS seemingly increasing with time, a future difference in fracture events cannot be ruled out.

This study's main strength is its randomized, controlled design with a follow-up time of 10 years of 129 patients with defined mild PHPT. The study also has limitations. Because DXA scans were obtained over a period of 19 years from several different machines, the optimal retrieval of raw data and analysis of all scans in one run with the same settings[19] was not possible. However, the BMD–*T*-score plot facilitated the identification of random and systematic differences and made it possible to correct for most of these. Missing DXA data is also a concern, but mixed model analysis of DXA data allowed all available data to be used statistically, even if the patient did not have data from all three time points. Drugs against osteoporosis were allowed in the study. During the study, 18 of the 129 patients used bisphosphonates at some time point (13 OBS/5 PTX), with a significant difference between groups ($p = 0.039$). Despite the fact that more patients in the OBS group received bisphosphonates in the 10-year cohort, a treatment effect of PTX compared with OBS on BMD was found. Likewise, estrogens were allowed but without difference in use between the groups. The 1991 NIH conference[3] recommended surgery for patients with creatinine clearance reduced by 30%, which in the pragmatic study protocol was defined for the age group to be reflected by a creatinine level >130 μmol/L. The initially relatively high drop-out rate is also a limitation. Surgically treated patients in the OBS group can be considered a limitation but also reflects a real-world situation with some patients meeting criteria for or wanting surgery at some stage after inclusion in the OBS group. This is also inherent to the modified intention-to-treat principle used for analyses in our study. Nevertheless, the per protocol analysis did not change the conclusions of the study. No power calculations were made for fractures, one of the secondary endpoints of the study. Thus, we cannot rule out a type two error due to the study potentially being underpowered as a cause of no detectable difference in fracture rates between groups.

In conclusion, PTX showed positive treatment effects on BMD for all analyzed compartments compared with OBS. However, BMD can be considered a surrogate endpoint, and we could not demonstrate a treatment effect of PTX on fractures in the 10-year cohort, neither with mITT analysis nor with per protocol analysis. Thus, from a bone health perspective, observation can

be considered a safe option in a 10-year perspective for the majority of patients with mild PHPT.

Conflicts of Interest

All authors state that they have no conflicts of interest.

Acknowledgments

The authors thank Viveca Åberg (Karolinska University Hospital, Stockholm, Sweden) and Jenny Tiberg (Sahlgrenska University Hospital, Gothenburg, Sweden) for their assistance throughout the study. This study has received support from Swedish Government grants under the ALF agreement (ALFGBG-592861, ALFGBG-789751, ALFGBG-434941, ALFGBG-719031), the Göteborg Medical Society (GLS-596301), and from the South-Eastern Norway Regional Health Authority (2018015).

Author Contributions

Karolina Lundstam: Data curation; formal analysis; funding acquisition; methodology; validation; visualization; writing – original draft; writing – review and editing. **Mikkel Pretorius:** Data curation; formal analysis; funding acquisition; methodology; validation; writing – review and editing. **Jens Bollerslev:** Conceptualization; funding acquisition; investigation; project administration; supervision; writing – review and editing. **Kristin Godang:** Data curation; investigation; project administration; supervision; validation; writing – review and editing. **Morten Wang Fagerland:** Formal analysis; methodology; writing – review and editing. **Charlotte Mollerup:** Investigation; writing – review and editing. **Stine Lyngvi Fougner:** Investigation; writing – review and editing. **Ylva Pernow:** Investigation; writing – review and editing. **Turid Aas:** Investigation; writing – review and editing. **Ola Hessman:** Investigation; writing – review and editing. **Thord Rosén:** Investigation; writing – review and editing. **Jörgen Nordenström:** Investigation; writing – review and editing. **Svante Jansson:** Investigation; writing – review and editing. **Mikael Hellström:** Funding acquisition; methodology; project administration; supervision; writing – review and editing. **Ansgar Heck:** Data curation; investigation; project administration; supervision; writing – review and editing.

Peer Review

The peer review history for this article is available at https://publons.com/publon/10.1002/jbmr.4763.

Data Availability Statement

Access to de-identified data from the SIPH-trial will be made available upon e-mail request to the corresponding author. Data will be shared within the Scandinavian laws and legislations.

References

1. Wilhelm SM, Wang TS, Ruan DT, et al. The American Association of Endocrine Surgeons Guidelines for definitive management of primary hyperparathyroidism. JAMA Surg. 2016;151(10):959-968.
2. Silverberg SJ, Clarke BL, Peacock M, et al. Current issues in the presentation of asymptomatic primary hyperparathyroidism: proceedings of the fourth international workshop. J Clin Endocrinol Metab. 2014;99(10):3580-3594.
3. NIH Conference. Diagnosis and management of asymptomatic primary hyperparathyroidism: consensus development conference statement. Ann Intern Med. 1991;114(7):593-597.
4. Bilezikian JP, Potts JT Jr, Fuleihan Gel H, et al. Summary statement from a workshop on asymptomatic primary hyperparathyroidism: a perspective for the 21st century. J Clin Endocrinol Metab. 2002;87(12):5353-5361.
5. Bilezikian JP, Khan AA, Potts JT Jr. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the third international workshop. J Clin Endocrinol Metab. 2009;94(2):335-339.
6. Bilezikian JP, Brandi ML, Eastell R, et al. Guidelines for the management of asymptomatic primary hyperparathyroidism: summary statement from the fourth international workshop. J Clin Endocrinol Metab. 2014;99(10):3561-3569.
7. Mosekilde L. Primary hyperparathyroidism and the skeleton. Clin Endocrinol (Oxf). 2008;69(1):1-19.
8. Hagstrom E, Lundgren E, Mallmin H, Rastad J, Hellman P. Positive effect of parathyroidectomy on bone mineral density in mild asymptomatic primary hyperparathyroidism. J Intern Med. 2006;259(2):191-198.
9. Cusano NE, Rubin MR, Silva BC, et al. Skeletal microstructure and estimated bone strength improve following parathyroidectomy in primary hyperparathyroidism. J Clin Endocrinol Metab. 2018;103(1):196-205.
10. Rolighed L, Vestergaard P, Heickendorff L, et al. BMD improvements after operation for primary hyperparathyroidism. Langenbecks Arch Surg. 2013;398(1):113-120.
11. Christiansen P, Steiniche T, Brixen K, et al. Primary hyperparathyroidism: short-term changes in bone remodeling and bone mineral density following parathyroidectomy. Bone. 1999;25(2):237-244.
12. Rubin MR, Bilezikian JP, McMahon DJ, et al. The natural history of primary hyperparathyroidism with or without parathyroid surgery after 15 years. J Clin Endocrinol Metab. 2008;93(9):3462-3470.
13. Rao DS, Phillips ER, Divine GW, Talpos GB. Randomized controlled clinical trial of surgery versus no surgery in patients with mild asymptomatic primary hyperparathyroidism. J Clin Endocrinol Metab. 2004;89(11):5415-5422.
14. Ambrogini E, Cetani F, Cianferotti L, et al. Surgery or surveillance for mild asymptomatic primary hyperparathyroidism: a prospective, randomized clinical trial. J Clin Endocrinol Metab. 2007;92(8):3114-3121.
15. Lundstam K, Heck A, Godang K, et al. Effect of surgery versus observation: skeletal 5-year outcomes in a randomized trial of patients with primary HPT (the SIPH study). J Bone Miner Res. 2017;32(9):1907-1914.
16. Lundstam K, Heck A, Mollerup C, et al. Effects of parathyroidectomy versus observation on the development of vertebral fractures in mild primary hyperparathyroidism. J Clin Endocrinol Metab. 2015;100(4):1359-1367.
17. Pretorius M, Lundstam K, Heck A, et al. Mortality and morbidity in mild primary hyperparathyroidism: results from a 10-year prospective randomized controlled trial of parathyroidectomy versus observation. Ann Intern Med. 2022;175(6):812-819.
18. Bollerslev J, Jansson S, Mollerup CL, et al. Medical observation, compared with parathyroidectomy, for asymptomatic primary hyperparathyroidism: a prospective, randomized trial. J Clin Endocrinol Metab. 2007;92(5):1687-1692.
19. Lundstam K, Godang K, Pretorius M, et al. The influence of DXA hardware, software, reference population and software analysis settings on the bone mineral density and T-score relationship. J Clin Densitom. 2022;25(1):24-33.
20. Silverberg SJ, Shane E, Jacobs TP, Siris E, Bilezikian JP. A 10-year prospective study of primary hyperparathyroidism with or without parathyroid surgery. N Engl J Med. 1999;341(17):1249-1255.
21. Nordenstrom E, Westerdahl J, Bergenfelz A. Recovery of bone mineral density in 126 patients after surgery for primary hyperparathyroidism. World J Surg. 2004;28(5):502-507.

22. Bolland MJ, Grey AB, Orr-Walker BJ, et al. Prospective 10-year study of postmenopausal women with asymptomatic primary hyperparathyroidism. N Z Med J. 2008;121(1277):18-29.

23. Castellano E, Attanasio R, Boriano A, et al. Sex difference in the clinical presentation of primary hyperparathyroidism: influence of menopausal status. J Clin Endocrinol Metab. 2017;102(11):4148-4152.

24. Dadon T, Tsvetov G, Levi S, Gorshtein A, Slutzky-Shraga I, Hirsch D. Gender differences in the presentation, course and outcomes of primary hyperparathyroidism. Maturitas. 2021;145:12-17.

25. Eller-Vainicher C, Filopanti M, Palmieri S, et al. Bone quality, as measured by trabecular bone score, in patients with primary hyperparathyroidism. Eur J Endocrinol. 2013;169(2):155-162.

26. Vignali E, Viccica G, Diacinti D, et al. Morphometric vertebral fractures in postmenopausal women with primary hyperparathyroidism. J Clin Endocrinol Metab. 2009;94(7):2306-2312.

27. Seib CD, Meng T, Suh I, et al. Risk of fracture among older adults with primary hyperparathyroidism receiving parathyroidectomy vs nonoperative management. JAMA Intern Med. 2022;182(1):10–18.

28. Curtis EM, Moon RJ, Harvey NC, Cooper C. The impact of fragility fracture and approaches to osteoporosis risk assessment worldwide. Bone. 2017;104:29-38.

29. Bergh C, Wennergren D, Möller M, Brisby H. Fracture incidence in adults in relation to age and gender: a study of 27,169 fractures in the Swedish fracture register in a well-defined catchment area. PLoS One. 2020;15(12):e0244291.

Baksidetekst

Primary Hyperparathyroidism (PHPT) is a common endocrine disorder characterized by elevated calcium levels and inappropriately high parathyroid hormone (PTH) levels. Mild PHPT, without apparent symptoms or signs, is now the predominant form, probably due to an increase in measurements of both calcium and PTH.

The overall aim of this thesis was to investigate if surgical removal of the parathyroid adenoma has an effect on long-term mortality and morbidity in mild PHPT.

In the Scandinavian Study of Primary Hyperparathyroidism (SIPH-study), we followed 191 patients, of whom 95 were randomized to parathyroidectomy (PTX), and 96 were randomized to observation (OBS). Both groups were followed with annual visits over a period of 10 years.

The study's primary end-point was mortality. The mortality was equal in both groups after ten years and after an extended follow-up period of up to twenty years.

We further investigated the incidence of cardiovascular and cerebrovascular events, fractures, malignancies, and kidney stones during the ten years of follow-up. There were no statistical differences between the groups for these secondary outcomes. Patient reported outcomes on quality of life indicated a subtle improvement in the PTX group. However, this improvement was weaker than what is generally accepted to be clinically relevant.

Lastly, we investigated the effect of PTX on bone mineral density with dual-energy X-ray absorptiometry (DXA). Here, we found a significant treatment effect in all the measured DXA compartments in favor of the PTX group. The OBS group had a significant decrease in all compartments compared to baseline. However, this difference did not transfer to an increased fracture risk over ten years.

In summary, mild PHPT seems to be a stable disease, and PTX does not appear to reduce mortality or morbidity over the course of a decade. Observation of patients with mild PHPT seems safe from a 10-year perspective.