Mortality Benefit of Ultrasound for Incidental Thyroid Nodules Identified with PET Imaging: A Non-Inferiority Emulated Target Trial

Protocol Number: 2024P002073-1

Principal Investigator: Jeffrey P. Guenette M.D. M.P.H.

Funded by: Agency for Healthcare Research and Quality

Version Number: v.1.0

20 November 2025

Summary of Changes from Previous Version:

Affected Section(s)	Summary of Revisions Made	Rationale

Table of Contents

STATEMENT OF COMPLIANCE

The emulated target trial will be carried out in accordance with International Conference on Harmonisation Good Clinical Practice (ICH GCP) and the following:

- United States (US) Code of Federal Regulations (CFR) applicable to clinical studies (45 CFR Part 46, 21 CFR Part 50, 21 CFR Part 56, 21 CFR Part 312, and/or 21 CFR Part 812)

National Institutes of Health (NIH)-funded investigators and clinical trial site staff who are responsible for the conduct, management, or oversight of NIH-funded clinical trials have completed Human Subjects Protection and ICH GCP Training.

This study has been exempted by the Mass General Brigham Institutional Review Board (IRB) for review and approval. The IRB protocol number is: 2024P002073.

1 PROTOCOL SUMMARY

1.1 SYNOPSIS

Title:	Mortality Benefit of Ultrasound for Incidental Thyroid Nodules Identified with PET Imaging: A Non-Inferiority Emulated Target Trial
Study Hypothesis:	We hypothesize that all-cause mortality in patients with an incidental thyroid nodule on PET-CT who did not have thyroid ultrasound within 3 months is no worse than 5% lower than those who did have ultrasound.
Outcome Measures:	Primary Outcome: All-cause mortality. Secondary Outcomes: Numbers of thyroid cancer diagnoses, thyroid ultrasounds, thyroid biopsies, and thyroid surgeries Exploratory Outcomes: Types of thyroid cancer diagnoses
Study Population:	All patients age 18-years and older with incidental thyroid nodule on PET-CT performed between 1/1/2015 and 12/31/2021.
Sites/Facilities:	Mass General Brigham healthcare system including Massachusetts General Hospital, Brigham and Women's Hospital, Mass Eye and Ear, and associated community sites.
Exposure:	Thyroid ultrasound evaluation within 3-months of PET-CT

1.2 TARGET VS. EMULATED TRIAL TABLE

	Target Clinical Trial	**Observational Emulation**
Design	Non-inferiority randomized controlled trial	Retrospective non-inferiority cohort study
Aim	Assess whether all-cause mortality in patients with an incidental thyroid nodule on PET-CT who did	Same

	not have thyroid ultrasound within 3 months is no worse than 5% lower than those who did have ultrasound	
Inclusion	Thyroid nodule on PET-CT Age 18 or older Clinic note within 36-months before the PET-CT	Same, with PET-CT scans between 1/1/2015 to 12/31/2021
Exclusion	Thyroid ultrasound in prior 3 years Prior thyroid cancer diagnosis	Same
Treatment Strategies	(1) Thyroid ultrasound ordered to be performed within 3 months (2) No thyroid ultrasound ordered	(1) Thyroid ultrasound performed within 3 months (2) No thyroid ultrasound performed within 3 months
Treatment Assignment	Randomized to treatment strategy	Non-randomized to treatment strategy. Randomization emulated via cloning patients in both arms.
Treatment Implementation	3-month grace period	3-month grace period
Outcome	All-cause mortality	Same
Type of Outcome	Failure time	Same
Follow-up Start	PET-CT date, equivalent to treatment assignment	PET-CT date, not equivalent to treatment assignment
Follow-up End	Earliest of death, 7-years from follow-up start, or study end	Same
Censoring	Administrative censoring	Administrative censoring Artificial censoring of clones
Causal Contrast	Intention to treat and per protocol	Per protocol effect only
Non-Inferiority Margin	5%	Same
Statistical Analysis	Non-inferiority intention to treat analysis that consists of comparing probability of mortality between treatment groups at 7 years.	Non-inferiority analysis using clone-censor-weight strategy: 1. Each subject is cloned at study baseline; one clone assigned to each treatment strategy.

	Kaplan-Meier to estimate probability of mortality in each arm Construct a 95% confidence interval for the absolute difference between the two arms, Δ_{RD} Taking the type I error rate, alpha, to be 0.05, non-inferiority is concluded if the upper limit of a two-sided 90% = 100(1-2alpha)% confidence interval for $\Delta_{RD}$is less than the prespecified margin.	2. Person-time for each clones is censored when a their observed treatment strategy is non-adherent. That is, clones assigned to undergo thyroid ultrasound within 3 months are censored at 3 months of follow-up if no ultrasound is performed. Similarly, clones assigned to not undergo an ultrasound within 3 months are censored at the time an ultrasound is performed. 3. A pooled logistic regression model is fit for the (conditional) probability of adherence, adjusting for a priori specified baseline factors that predict adherence. 4. A marginal structural pooled logistic regression model is fit to the mortality outcome data, as a function of time and treatment (with interaction), using inverse-probability of censoring weights from step 3. 5. Treatment-specific marginal probabilities of all-cause mortality are estimated from the fit in step 4, from which the causal risk difference, Δ_{RD}, is estimated along with a two-sided 90% = 100(1-2alpha)% confidence interval (constructed via the bootstrap). 6. Taking the type I error rate, alpha, to be 0.05, non-inferiority is concluded if the upper limit of a two-sided 90% = 100(1-2alpha)% confidence interval for $\Delta_{RD}$is less than the prespecified margin.
Secondary Outcomes	Cumulative over 7-years from PET: Number of thyroid cancers Number of thyroid ultrasounds Number of thyroid biopsies Number of thyroid surgeries	Same
Exploratory Analysis	Types of thyroid cancer diagnoses	Same

2 INTRODUCTION

2.1 STUDY RATIONALE

Early pre-symptomatic detection of new or recurrent malignancy is frequently presumed to identify disease at a more treatable stage and, thus, result in better survival than detection due to symptoms.[1] However, evidence often only equivocally supports and frequently contradicts this presumption.[2–7] Furthermore, most clinical providers incorrectly estimate the harm versus benefit tradeoffs inherent in pre-symptomatic detection strategies.[1,8–14] Based on the background information provided in Section 2.2, we hypothesize that all-cause mortality in patients with an incidental thyroid nodule on PET-CT who did not have thyroid ultrasound within 3 months is no worse than 5% lower than those who did have ultrasound.

2.2 BACKGROUND

A recent multi-decade SEER study concluded that overdiagnosis of thyroid cancer is a crucial unresolved public health issue.[15] Overuse of thyroid ultrasound, particularly for clinically unsupported reasons, is a major driver of the overdiagnosis and overtreatment of small non-palpable indolent thyroid cancers.[16] Although evidence is not currently available, it is likely that ultrasound evaluation of incidental thyroid nodules on PET examinations fall into this overuse category, particularly given that the baseline 10-year mortality of patients undergoing PET examinations at our institution are in the range of 30%-50%.

Briefly, thyroid nodules have been detected in up to 65% of people in the United States (U.S.) at autopsy with most autopsy series showing a prevalence of ~50%.[17] However, only 1.1% of people will be diagnosed with thyroid cancer[18–20] and only 0.4% of all cancer-related deaths are attributable to thyroid cancer.[20] The 5-year survival for all patients with thyroid cancer is 98.6%[21] and 30% of deaths are attributable to anaplastic thyroid carcinoma,[22] which accounts for only 1% of thyroid cancers.[21]

Surgical complication rates in thyroidectomy include a 12% rate of intermediate- to long-term thyroid surgery-specific complications including hypothyroidism, hypoparathyroidism, and hypocalcemia from 1-month to 1-year following surgery in addition to a 6% rate of postoperative fever, hemorrhage, emergency intubation, tracheostomy, pneumonia, and cardiopulmonary and thrombotic complications within 30 days of surgery.[23]

Current American College of Radiology guidelines recommend thyroid ultrasound evaluation of incidental nodules identified on PET with determination for biopsy based on ultrasound characteristics.[24] Based on the TI-RADS lexicon for reporting thyroid ultrasound,[25] 26% of thyroid nodules are recommended for biopsy.[26]

Given the high incidence of thyroid nodules, low rates of thyroid cancer, and low associated mortality, it is likely that workup of incidental thyroid nodules does not impact mortality yet results in patient harm. We thus aim to provide evidence to help understand whether ultrasound evaluation of incidental thyroid nodules on PET examinations contributes to overdiagnosis and overtreatment.

2.3 RISK/BENEFIT ASSESSMENT

Known potential risk to subjects is limited to inadvertent breach of confidentiality due to exposure of identified data. This risk will be minimized by following data safety criteria outlined in Section 9. There

are likely no direct benefits to subjects, but it is hoped that information gathered from this study may be helpful to the medical care of patients in the future.

3 OBJECTIVES AND ENDPOINTS

PRIMARY PURPOSE (OBJECTIVES)	OUTCOME MEASURES	JUSTIFICATION FOR OUTCOME MEASURES
Primary		
Estimate whether all-cause mortality in patients with an incidental thyroid nodule on PET-CT who did not have thyroid ultrasound within 3 months was non-inferior within a 5% margin to those who did have ultrasound.	All-cause mortality	Determined by three thyroid endocrinologists and a head and neck surgeon to be a reasonable timeframe in which thyroid cancers have time to develop and contribute to mortality but not so much that it becomes difficult to separate thyroid cancer-related mortality from other causes.
Secondary		
Estimate whether numbers of thyroid cancers, thyroid ultrasounds, thyroid biopsies, and thyroid surgeries are higher in patients with an incidental thyroid nodule on PET-CT who had versus did not have thyroid ultrasound within 3 months.	Number of thyroid cancers, number of thyroid ultrasounds, number of thyroid biopsies, number of thyroid surgeries	Between-group differences in these outcomes will help establish whether the estimated non-inferiority result is clinically meaningful and justifiable due to increased morbidity in those who under further thyroid nodule characterization.
Tertiary/Exploratory		
Explore whether types of thyroid cancer diagnoses are different among patients with an incidental thyroid nodule on PET-CT who had versus did not have thyroid ultrasound within 3 months.	Thyroid cancer subtypes	It is likely that mostly indolent papillary thyroid carcinomas are identified when incidental thyroid nodules are further evaluated.

4 STUDY DESIGN

4.1 OVERALL DESIGN

We hypothesize that all-cause mortality in patients with an incidental thyroid nodule on PET-CT who did not have thyroid ultrasound (the exposure) within 3 months of the PET-CT is non-inferior within a 5% margin to those who have thyroid ultrasound at 7-years. That is, among patients with an incidental

thyroid nodule on PET-CT, mortality is no more than 5% larger (in absolute difference) for those who do not have thyroid ultrasound compared to those who do. We will also report mortality differences at landmark timeframes of 1-year, 3-years, 5-years, and 10-years.

Understanding an expected 30%-50% baseline 10-year mortality rate in patients who undergo PET imaging at our institution, a 5% non-inferiority margin was chosen based on the most conservative proposed margin by 2 thyroid endocrinologists (one did not respond) and a head and neck surgeon (other proposed margins were 7% and 10%). This margin decision was informed by typical margins of 5% to 10% used in non-inferiority trials of medications with outcomes involving mortality[27] and by recent non-inferiority trials of thyroid cancer treatments.[28,29] The observational nature of this non-inferiority study and absence of prior studies that show effect sizes for the reference make ideal margin point-estimate and fixed-estimate calculations impossible, but our expert panel method is supported as a practical necessity[30] to provide guidance on the fraction of the point estimate of the active comparator that stakeholders are willing to lose at a cost of gaining other advantages (in this case, likely fewer biopsies and surgeries), as in the point-estimate margin calculation method.[31] Given this careful pre-specification of our margin with this panel decision, we will follow published methodology recommendations[30] and consider re-evaluation of the margin based on the secondary benefits seen in the emulated trial or we may conduct the final analysis without reference to a margin. These decisions will be clearly detailed in final manuscripts.

To estimate group differences in mortality, we will conduct a non-inferiority emulated target trial utilizing clone-censor weighting to address potential immortal time bias introduced by the 3-month grace period. We will adjust for all variables labeled as demographic, potential confounder, or mortality risk adjustor in as listed in the Section 7, below.

We will stratify analyses based on baseline disease severity (estimated 5-year relative survival risk) and disease status (progression, lymph node involvement, other sites of metastases).

All subjects will be accrued from the Mass General Brigham healthcare system, which includes two academic medical centers, a specialty head and neck hospital, and multiple community hospitals and numerous community clinics.

4.2 SCIENTIFIC RATIONALE FOR STUDY DESIGN

The baseline 10-year mortality of patients undergoing PET examinations at our institution are in the range of 30%-50%. Our interest is in whether, following the detection of a thyroid nodule, doing nothing (i.e. no thyroid ultrasound) results in mortality that is no worse than 5% more than doing something (i.e. thyroid ultrasound). Thus we are using a non-inferiority design. Since performing an ultrasound aligns with ACR recommendations, we've designated the imaged group as the reference group.

4.3 END OF STUDY DEFINITION

Mortality data collected up to October 15, 2025. Patients included in the study until they died or until administratively censored.

5 STUDY POPULATION

5.1 INCLUSION CRITERIA

In order to be included in this study, an individual must meet all of the following criteria:

1. Age >=18
2. Thyroid nodule on PET-CT performed 1/1/2015 to 12/31/2021
3. At least one clinical note in the EHR from the 36-month window prior to the PET

Note, there are no restrictions on the indication for the index PET-CT.

5.2 EXCLUSION CRITERIA

An individual who meets any of the following criteria will be excluded:

1. Thyroid ultrasound listed in the medical record in the prior 3 years
2. Documented history of prior thyroid cancer diagnosis

6 STUDY EXPOSURE

6.1 STUDY EXPOSURE DEFINITION

For this target trial emulation, exposure is whether a thyroid ultrasound is performed within 3 months of the index PET-CT.

6.2 MEASURES TO MINIMIZE BIAS

Data extraction from the electronic health record data warehouse using SQL and Python scripts will necessarily be conducted with patient identifiers. All data will be anonymized prior to analysis and all blinded data utilized for analysis will be published publicly as utilized in the analysis. Study group assignments will be strictly performed based on whether or not thyroid ultrasound was performed within the 3-month period following the PET-CT; there is no subjectivity to the assignment.

6.3 LOST TO FOLLOW-UP

Complete mortality data is available in the electronic health record via regular updates from electronic death registry systems, regardless of patients' continued care within the study healthcare system; so there is no loss to follow-up for the primary outcome. The 3-month grace period between the PET examination and the ultrasound will be addressed with clone-censor weighting. For the secondary outcomes, we will consider patients who did not have any provider note for 2-years following the PET-CT to be lost to follow-up.

7 STUDY PROCEDURES

All of the following variables, except the last 12 disease status variables, will be extracted from the electronic health system research data warehouse using Snowflake, SQL scripts, and Python scripts. The extracted data will be coded as described in R. The last 12 disease status variables will be extracted and coded from the PET-CT reports by an instance of ChatGPT4 in Azure that is secure behind the healthcare system firewall. Data will be included for patients who had PET-CT examinations performed between

1/1/2015 and 12/31/2021, which allows a 3-year period for complete propagation of death records to our healthcare system's electronic health records. All HIPAA rules and local institutional and IRB requirements will be followed for data collection.

Exposure Variable	**Explanation**
thyroidus3mospost	Exposure: 1 = thyroid ultrasound performed within 3-months following imaging examination; 0 = thyroid US not performed within 3-months following imaging examination

Outcome Variables	**Explanation**
mortality_binary	Primary Outcome Binary: 1 = died within follow-up period; 0 = did not die within follow-up period
follow_up_in_days	Primary Outcome Continuous: Number of days from imaging examination to death, 10-years, or end of study, whichever comes first
thyroid_cancer_dx_post_pet	Secondary Outome Binary: 1 = ICD10 C73 diagnosis after PET exam; 0 = no ICD10 C73 diagnosis
post_pet_thyroid_excision	Secondary Outcome Binary: 1 = CPT code 60100 for thyroid excision performed after PET exam
thyroid_ultrasounds	Secondary Outcome Continuous: Number of thyroid ultrasound exams performed after PET exam
thyroid_biopsies	Secondary Outcome Continuous: Number of thyroid biopsies performed after PET exam
thyroid_cancer_subtype	Exploratory Outcome: ICD10 C73 subcategory

Inclusion/Exclusion Variables	**Explanation**
thyroid_nodule	Inclusion Criteria: 1 = thyroid nodule reported in imaging examination report; 0 = thyroid nodule not reported
priorcareinsystem	Inclusion Criteria: 1 = clinic note within 36-months prior to PET; 0 = no clinic note in that timeframe
age	Inclusion Criteria and Demographic: age in years at time of imaging examination
priorthyroidus	Exclusion Criteria: 1 = thyroid ultrasound performed within 3-years prior to imaging examination; 0 = thyroid US not performed within 3-years prior to imaging examination
priorthyroidcancer	Exclusion Criteria: ICD-10 C73 or ICD-9 193 before PET
NoSustainedCare	Exclusion Criteria: 1 = no provider note within system for 2-years following recommendation

Examination Variables	**Explanation**
exam_id	Identifier: Anonymized imaging examination ID
patient_id	Identifier: Anonymized patient ID
exam_year	Year imaging examination was performed

pet_cat	Potential Confounder: PET Type - 1= "NM PET WHOLE BODY OUTSIDE WITH INTERPRETATION OR CONSULT"; 2="NM PET CT SKULL BASE TO MID THIGHS", "NM PET CT SKULL BASE TO MID THIGH WITH DEDICATED BRAIN", "NM PET CT SCALP TO TOES"; 3="NM PET CT NEUROENDOCRINE TUMOR LOCALIZATION"; 4 = "NM PET CT PROSTATE CANCER IMAGING"; 5= "NM PET CT BONE SCAN SODIUM FLUORIDE"
exam_care_level	Potential Confounder: care level at time of examination - E = Emergency; I = Inpatient; O = Outpatient
ReferringProviderDepartment	Potential Confounder: specialty of provider who ordering imaging examination; specialties categorized into 6 broader groups: surgery, pediatrics, emergency medicine,neurology, oncology, other medicine, and missing
exam_site	Potential Random Effect: anonymized location of scanner where examination performed

Additional Patient Variables	**Explanation**
sex	Demographic: 1 = male; 2 = female; 3 = other/unknown
adi	Demographic: national area deprivation index on scale 0 (low deprivation) to 100 (high deprivation) based on listed home address in 2025
coi	Demographic: national childhood opportunity index on scale 0 (low opportunity) to 100 (high opportunity) based on listed home address in 2025
svi	Demographic: social vulnerability index on scale 0 (low vulnerability) to 100 (high vulnerability) based on listed home address in 2025
language	Demographic: self-reported language
race	Demographic: self-reported race
ethnicity	Demographic: self-reported ethnicity
smoking	Potential Confounder: 0 = never; 1 = former; 2 = active; No_self-report = not reported; most recently recorded prior to imaging
alcohol	Potential Confounder: 0 = never; 1 = occasional; 2 = heavy; most recently recorded prior to imaging
social_hx_contact	Date social history was taken
socialhx_days_from_exam	Days from social history to PET exam date (amount is negative if social history was taken after)
obesity	Potential Confounder: 1 = diagnosis of obesity (ICD-10 E66, ICD-9 278.00 or 278.01) at time of imaging examination OR OR cacluated BMI >30 at timepoint within 1 year before and closest to imaging examination; 0 = no diagnosis of obesity or BMI >30
metabolic_syndrome	Potential Confounder: 1 = diagnosis of metabolic syndrome (ICD-10 , ICD-9 277.7) at time of imaging examination; 0 = no diagnosis of metabolic syndrome at time of imaging examination
mi	Mortality Risk Adjustor: Myocardial Infarction: ICD-10 I21.9; ICD-9 410; recorded prior to date of imaging examination

chf	Mortality Risk Adjustor: Congestive Heart Failure: ICD-10 I50; ICD-9 428; recorded prior to date of imaging examination
pvd	Mortality Risk Adjustor: Peripheral Vascular Disease: ICD-10 I70, I73; ICD-9 443, 440.2, 440.3, 440.4; recorded prior to date of imaging examination
cva	Mortality Risk Adjustor: Cerebrovascular Accident: ICD-10 I60-I63, G45; ICD-9 430-438; recorded prior to date of imaging examination
dementia	Mortality Risk Adjustor: Dementia: ICD-10 F01-F03; ICD-9 290; recorded prior to date of imaging examination
cpd	Mortality Risk Adjustor: Chronic Pulmonary Disease: ICD-10 J42-J44; ICD-9 491-492; recorded prior to date of imaging examination
ctd	Mortality Risk Adjustor: Connective Tissue Disease: ICD-10 M30-M36; ICD-9 710; recorded prior to date of imaging examination
pud	Mortality Risk Adjustor: Peptic Ulcer Disease: ICD-10 K25-K28; ICD-9 531-534; recorded prior to date of imaging examination
ld	Mortality Risk Adjustor: Liver Disease: ICD-10 K70-K77; ICD-9 570-573; recorded prior to date of imaging examination
dm	Mortality Risk Adjustor: Diabetes Millitus: ICD-10 E8-E13; ICD-9 249-250; recorded prior to date of imaging examination
hemiplegia	Mortality Risk Adjustor: Hemiplegia: ICD-10 G81; ICD-9 342; recorded prior to date of imaging examination
ckd	Mortality Risk Adjustor: Chronic Kidney Disease: ICD-10 N18; ICD-9 585; recorded prior to date of imaging examination
cancer	Mortality Risk Adjustor: Cancer: ICD-10 C00-C96; ICD-9 140-209; recorded prior to date of imaging examination
obesity_icd	Mortality Risk Adjustor: Diagnosis of obesity (ICD-10 E66, ICD-9 278.00 or 278.01) at time of imaging examination
bmi	Mortality Risk Adjustor: Body mass index at time of imaging examination
hncancer_binary	Mortality Risk Adjustor: Head/Neck Cancer: ICD-10 C00-C14; ICD-9 140-149; recorded prior to date of imaging examination
radiation_binary	Mortality Risk Adjustor: Radiation Therapy: CPT 77401-77416, G6003-G6014,77385-77386,G6015-G6016
priorneckradiation	Mortality Risk Adjustor: Prior head/neck cancer radiation therapy (hncancer_binary =1 and radiation_binary =1 , then priorneck radiation =1; otherwise, 0)
Fiveyr_atPET	Mortality Risk Adjustor: 5 year survival probability associated with cancer dx received before PET exam (if multiple cancers before PET, lowest survival probability)
DiseaseFree	Potential Confounder: 1 = PET-CT report indicates there is no active malignancy
DiseaseProgression	Potential Confounder: 1= PET-CT report indicates disease progression
LymphNodes	Potential Confounder: 1= PET-CT report indicates there are lymph node metastases
NonNodeSites	Potential Confounder: 1= PET-CT report indicates that there are axillary lymph node metastases

AxillaryNodes	Potential Confounder: 1= PET-CT report indicates that there are neck or supraclavicular lymph node metastases
MediastinalNodes	Potential Confounder: 1= PET-CT report indicates that there are mediastinal (including paratracheal, paraesophageal, prevascular, subcarinal, perihilar, or hilar) lymph node metastases
AbdominalNodes	Potential Confounder: 1= PET-CT report indicates that there are abdominal (including peri-portal, celiac, mesenteric, retroperitoneal, iliac) lymph node metastases
InguinalNodes	Potential Confounder: 1= PET-CT report indicates that there are inguinal lymph node metastases
NonNodeSites	Potential Confounder: 1= PET-CT report indicates that there are metastases outside of the lymph nodes (liver, lungs, bones, soft tissue, stomach, bowel, colon, spleen, pancreas, adrenal glands, or other organs other than the thyroid and other than the primary site)
Lung	Potential Confounder: 1= PET-CT report indicate that there are likely metastatic lesions in the lungs
Bone	Potential Confounder: 1= PET-CT report indicates that there are likely metastatic lesions in the bones
Liver	Potential Confounder: 1= PET-CT report indicate that there are likely metastatic lesions in the liver

8 STATISTICAL CONSIDERATIONS

8.1 STATISTICAL HYPOTHESES

Primary Outcome Measure

Hypothesis: All-cause mortality in patients with an incidental thyroid nodule on PET-CT who did not have thyroid ultrasound within 3 months is no worse than 5% lower than those who did have ultrasound within 3 months.

Null hypothesis: All-cause mortality in patients with an incidental thyroid nodule on PET-CT who did not have thyroid ultrasound within 3 months is more than 5% higher than those who did have ultrasound within 3 months.

Secondary Outcome Measure(s)

Hypotheses: (1) number of thyroid cancer diagnoses, (2) number of thyroid ultrasounds, (3) number of thyroid biopsies, and (4) number of thyroid surgeries are each higher in patients with an incidental thyroid nodule on PET-CT who had versus did not have thyroid ultrasound within 3 months.

Null hypotheses: (1) number of thyroid cancer diagnoses, (2) number of thyroid ultrasounds, (3) number of thyroid biopsies, and (4) number of thyroid surgeries are each not higher in patients with an incidental thyroid nodule on PET-CT who had versus did not have thyroid ultrasound within 3 months.

8.2 SAMPLE SIZE DETERMINATION

Informed by a preliminary sample of 84,775 patients undergoing CT and MRI imaging in 2015 with baseline rates of 10-year mortality of 39.5%, covariate adjusted power was assessed via simulation. For each simulated dataset, we resampled sex and age from the empirical covariate distribution among the sample of 84,775 patients and generated a binary indicator of 5- or 10-year mortality from a logistic regression model whose coefficients reflected the actual strength of association between each covariate and mortality. A logistic regression model was then fit to each simulated data set, where the primary coefficient of interest reflected the log odds ratio between treatment and mortality. Covariate adjusted power was calculated as the fraction of simulated datasets for which the two-sided confidence interval for this coefficient did not contain 0.2275, the value of the log-odds ratio corresponding to a 5% non-inferiority margin given the baseline mortality rate of 39.5%. Given that more covariates will be available once final analysis datasets are compiled, this method reflects a conservative estimate. Based on the results we anticipate having at least 75% power for a 5% non-inferiority margin.

8.3 POPULATIONS FOR ANALYSES

Single dataset for per-protocol analysis including all patients who met the inclusion/exclusion criteria.

8.4 STATISTICAL ANALYSES

8.4.1 GENERAL APPROACH

For descriptive statistics, categorical data will be presented, stratified by observed treatment status, as percentages while continuous data will be presented as medians and interquartile ranges. For non-inferiority (primary aim) inferential tests, statistical significance is defined as the estimated 7-year mortality two-sided confidence interval of the study group not crossing an absolute 5% lower 7-year mortality rate of the reference group. For superiority (secondary aims) inferential tests, statistical significance is defined as the estimated between-group difference two-sided confidence interval not crossing zero.

8.4.2 BASELINE DESCRIPTIVE STATISTICS

The study and reference groups will be compared on baseline characteristics, including demographics and disease status, using descriptive statistics without inferential statistics (there are no associated hypotheses).

8.4.3 ANALYSIS OF THE PRIMARY OUTCOME MEASURE(S)

The statistical analysis plan is outlined in detail in the Target Vs. Emulated Trial Table in Section 1.2. We do not anticipate substantive missing data.

8.4.4 ANALYSIS OF THE SECONDARY OUTCOME MEASURE(S)

No secondary endpoints are dependent on the findings from the primary endpoint. Analysis methods will follow the general approach for the primary outcome. Specifically, steps 1-4 of the Target Vs. Emulated Trial Table in Section 1.2 produce weights which reflect probability of ultrasound adherence. These weights will then be used to fit inverse probability weighted outcome models, which will either be

logistic regression or linear regression models, depending on the exact outcome data type. Correspondingly, results will be presented as risk differences or odds ratios with appropriate 95% confidence intervals.

8.4.5 SUB-GROUP ANALYSES

Stratified analyses will be performed based on baseline 5-year relative survival probabilities, calculated on the basis of the known cancer diagnoses and disease status at the time of PET.

8.4.6 EXPLORATORY ANALYSES

Descriptive statistics with counts and proportions of thyroid cancer subtypes in the study and reference groups will be performed.

9 SUPPORTING DOCUMENTATION AND OPERATIONAL CONSIDERATIONS

9.1 REGULATORY, ETHICAL, AND STUDY OVERSIGHT CONSIDERATIONS

9.1.1 INFORMED CONSENT PROCESS

Informed consent was waived by the institutional review board.

9.1.2 CONFIDENTIALITY AND PRIVACY

This study is a chart/imaging review only. There will be no interaction with the patients outside of the standard care being provided. Inclusion in the study will not impact or influence medical care in any way. Identifiable health information will be stored on a computer on the Mass General Brigham network with password protections enabled and anti-virus software or an encrypted laptop, with access to data limited to study staff. Direct identifiers, such as name and medical record number, will be removed once all of the data is collected and analysis performed on de-identified data.

9.1.3 KEY ROLES AND STUDY GOVERNANCE

Principal Investigator	
Name, degree, title	Jeffrey P. Guenette M.D. M.P.H.
Institution Name	Brigham and Women's Hospital
Address	1670 Tremont Street, 3rd Floor, Boston, MA
Phone Number	617-732-7260
Email	jpguenette@bwh.harvard.edu

9.1.4 DATA INTEGRITY, HANDLING, AND RECORD KEEPING

The principal investigator will verify that the study is conducted and data are generated, documented (recorded), and reported in compliance with the protocol, International Conference on Harmonisation

Good Clinical Practice (ICH GCP), and applicable regulatory requirements (e.g., Good Laboratory Practices (GLP), Good Manufacturing Practices (GMP)).

Data collection is the responsibility of the study staff under the supervision principal investigator. The principal investigator is responsible for ensuring the accuracy, completeness, legibility, and timeliness of the data reported.

Clinical data will be extracted from the electronic health records system and stored in csv files in restricted, password-protected, cloud folders behind the Mass General Brigham firewall.

Data retention will follow Mass General Brigham and NIH guidelines. Deidentified data will be publicly shared on Harvard Dataverse.

9.1.5 PROTOCOL DEVIATIONS

A protocol deviation is any noncompliance with the clinical trial protocol, International Conference on Harmonisation Good Clinical Practice (ICH GCP), or Manual of Procedures (MOP) requirements. The noncompliance may be either on the part of the participant, the investigator, or the study site staff. As a result of deviations, corrective actions are to be developed by the site and implemented promptly.

These practices are consistent with ICH GCP:

- 4.5 Compliance with Protocol, sections 4.5.1, 4.5.2, and 4.5.3
- 5.1 Quality Assurance and Quality Control, section 5.1.1
- 5.20 Noncompliance, sections 5.20.1, and 5.20.2.

It is the responsibility of the site investigator to use continuous vigilance to identify and report deviations to the reviewing Institutional Review Board (IRB) per their policies. The site investigator is responsible for knowing and adhering to the reviewing IRB requirements.

9.1.6 PUBLICATION AND DATA SHARING POLICY

This study will be conducted in accordance with the following publication and data sharing policies and regulations:

National Institutes of Health (NIH) Public Access Policy, which ensures that the public has access to the published results of NIH funded research. It requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funds to the digital archive PubMed Central upon acceptance for publication.

This study will comply with the NIH Data Sharing Policy. This study will be registered at ClinicalTrials.gov, and results information from this trial will be submitted to ClinicalTrials.gov. In addition, every attempt will be made to publish results in peer-reviewed journals. Data from this study may be requested from other researchers by reviewing the data repository on Harvard Dataverse.

9.1.7 CONFLICT OF INTEREST POLICY

Any actual conflict of interest of persons who have a role in the design, conduct, analysis, publication, or any aspect of this study will be disclosed and managed. Furthermore, persons who have a perceived conflict of interest will be required to have such conflicts managed in a way that is appropriate to their participation in the design and conduct of this study.

9.2 ADDITIONAL CONSIDERATIONS

None.

9.3 PROTOCOL AMENDMENT HISTORY

Version	Date	Description of Change	Brief Rationale
1.0	11/18/25	Initial Protocol	Not applicable

10 REFERENCES

1. Baum M. Harms from breast cancer screening outweigh benefits if death caused by treatment is included. British Medical Journal Publishing Group; 2013 Jan 23 [cited 2025 May 20]; Available from: https://www.bmj.com/content/346/bmj.f385.full

2. Monticciolo DL, Malak SF, Friedewald SM, Eby PR, Newell MS, Moy L, Destounis S, Leung JWT, Hendrick RE, Smetherman D. Breast Cancer Screening Recommendations Inclusive of All Women at Average Risk: Update from the ACR and Society of Breast Imaging. Journal of the American College of Radiology. Elsevier; 2021 Sept 1;18(9):1280–1288. PMID: 34154984

3. US Preventive Services Task Force. Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement. JAMA. 2024 June 11;331(22):1918–1930.

4. Statement on USPSTF Breast Cancer Screening Recommendations [Internet]. [cited 2025 May 5]. Available from: https://www.acr.org/News-and-Publications/Media-Center/2024/ACR-statement-on-final-USPSTF-breast-cancer-screening-recommendations

5. Dieng M, Lord SJ, Turner RM, Nieweg OE, Menzies AM, Saw RPM, Einstein AJ, Emmett L, Thompson JF, Lo SN, Morton RL. The Impact of Surveillance Imaging Frequency on the Detection of Distant Disease for Patients with Resected Stage III Melanoma. Ann Surg Oncol. 2022 May;29(5):2871–2881. PMCID: PMC8990943

6. Heiden BT, Eaton DB, Chang SH, Yan Y, Schoen MW, Thomas TS, Patel MR, Kreisel D, Nava RG, Meyers BF, Kozower BD, Puri V. Association between imaging surveillance frequency and outcomes following surgical treatment of early-stage lung cancer. J Natl Cancer Inst. 2023 Mar 9;115(3):303–310. PMCID: PMC9996218

7. Tzeng CWD, Abbott DE, Cantor SB, Fleming JB, Lee JE, Pisters PWT, Varadhachary GR, Abbruzzese JL, Wolff RA, Ahmad SA, Katz MHG. Frequency and Intensity of Postoperative Surveillance After Curative Treatment of Pancreatic Cancer: A Cost-Effectiveness Analysis. Ann Surg Oncol. 2013 July 1;20(7):2197–2203.

8. Zeng L, Helsingen LM, Nampo FK, Wang Y, Yao L, Siemieniuk RA, Bretthauer M, Guyatt GH. How do cancer screening guidelines trade off benefits versus harms and burdens of screening? A systematic survey. British Medical Journal Publishing Group; 2020 Dec 1 [cited 2025 May 20]; Available from: https://bmjopen.bmj.com/content/10/12/e038322.abstract

9. Dans LF, Silvestre MAA, Dans AL. Trade-off between benefit and harm is crucial in health screening recommendations. Part I: General principles. Journal of Clinical Epidemiology. 2011 Mar 1;64(3):231–239.

10. Silvestre MAA, Dans LF, Dans AL. Trade-off between benefit and harm is crucial in health screening recommendations. Part II: Evidence summaries. Journal of Clinical Epidemiology. 2011 Mar 1;64(3):240–249.

11. Germino JC, Elmore JG, Carlos RC, Lee CI. Imaging-based screening: maximizing benefits and minimizing harms. Clinical Imaging. 2016 Mar 1;40(2):339–343.

12. Hoffmann TC, Del Mar C. Clinicians' Expectations of the Benefits and Harms of Treatments, Screening, and Tests: A Systematic Review. JAMA Internal Medicine. 2017 Mar 1;177(3):407–419.

13. Heleno B, Thomsen MF, Rodrigues DS, Jørgensen KJ, Brodersen J. Quantification of harms in cancer screening trials: literature review. British Medical Journal Publishing Group; 2013 Sept 16 [cited 2025 May 20]; Available from: https://www.bmj.com/content/347/bmj.f5334.short

14. Caverly TJ, Hayward RA, Reamer E, Zikmund-Fisher BJ, Connochie D, Heisler M, Fagerlin A. Presentation of Benefits and Harms in US Cancer Screening and Prevention Guidelines: Systematic Review. JNCI: Journal of the National Cancer Institute. 2016 June 1;108(6):djv436.

15. Chen MM, Luu M, Sacks WL, Orloff L, Wallner LP, Clair JMS, Pitt SC, Ho AS, Zumsteg ZS. Trends in incidence, metastasis, and mortality from thyroid cancer in the USA from 1975 to 2019: a population-based study of age, period, and cohort effects. Lancet Diabetes Endocrinol. 2025 Mar;13(3):188–195. PMID: 39922210

16. Chen DW, Haymart MR. Unravelling the rise in thyroid cancer incidence and addressing overdiagnosis. Nat Rev Endocrinol. Nature Publishing Group; 2025 Sept 3;1–11.

17. Dean DS, Gharib H. Epidemiology of thyroid nodules. Best Practice & Research Clinical Endocrinology & Metabolism. 2008 Dec 1;22(6):901–911.

18. Smith-Bindman R, Lebda P, Feldstein VA, Sellami D, Goldstein RB, Brasic N, Jin C, Kornak J. Risk of thyroid cancer based on thyroid ultrasound imaging characteristics: results of a population-based study. JAMA Intern Med. 2013 Oct 28;173(19):1788–1796. PMCID: PMC3936789

19. Grussendorf M, Ruschenburg I, Brabant G. Malignancy rates in thyroid nodules: a long-term cohort study of 17,592 patients. Eur Thyroid J. 2022 May 30;11(4):e220027. PMCID: PMC9254276

20. Cancer of the Thyroid - Cancer Stat Facts [Internet]. SEER. [cited 2020 Nov 23]. Available from: https://seer.cancer.gov/statfacts/html/thyro.html

21. Kitahara CM, Schneider AB. Cancer Progress and Priorities: Epidemiology of Thyroid Cancer. Cancer Epidemiol Biomarkers Prev. 2022 July 1;31(7):1284–1297. PMCID: PMC9473679

22. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974-2013. JAMA. 2017 Apr 4;317(13):1338–1348.

23. Papaleontiou M, Hughes DT, Guo C, Banerjee M, Haymart MR. Population-Based Assessment of Complications Following Surgery for Thyroid Cancer. J Clin Endocrinol Metab. 2017 Apr 28;102(7):2543–2551. PMCID: PMC5505192

24. Hoang JK, Langer JE, Middleton WD, Wu CC, Hammers LW, Cronan JJ, Tessler FN, Grant EG, Berland LL. Managing Incidental Thyroid Nodules Detected on Imaging: White Paper of the ACR Incidental Thyroid Findings Committee. Journal of the American College of Radiology. 2015 Feb 1;12(2):143–150.

25. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, Hammers LW, Hamper UM, Langer JE, Reading CC, Scoutt LM, Stavros AT. ACR

Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. Journal of the American College of Radiology. Elsevier; 2017 May 1;14(5):587–595. PMID: 28372962

26. Hoang JK, Middleton WD, Tessler FN. Update on ACR TI-RADS: Successes, Challenges, and Future Directions, From the AJR Special Series on Radiology Reporting and Data Systems. AJR Am J Roentgenol. 2021 Mar;216(3):570–578. PMID: 33112199

27. Pong S, Urner M, Fowler RA, Mitsakakis N, Seto W, Hutchison JS, Science M, Daneman N. Testing for non-inferior mortality: a systematic review of non-inferiority margin sizes and trial characteristics. BMJ Open. British Medical Journal Publishing Group; 2021 Apr 1;11(4):e044480. PMID: 33879485

28. Mallick U, Newbold K, Beasley M, Garcez K, Wadsley J, Johnson SJ, Stephenson T, Gaze M, Goodman A, Jefferies S, Sivabalasingham S, Slevin N, Wilkinson DP, Macias-Fernandez E, Power D, Roques T, Speed L, Nutting C, Mochloulis G, Gerrard G, Candish C, Morgan S, Tripathi D, Truran P, Arthur C, Wieczorek A, Madhavan K, Maclean J, Boote D, Kim D, Pascoe A, Pitiyage G, Forsyth S, Ambrose E, Chang E, Farnell K, Hackshaw A. Thyroidectomy with or without postoperative radioiodine for patients with low-risk differentiated thyroid cancer in the UK (IoN): a randomised, multicentre, non-inferiority trial. The Lancet. Elsevier; 2025 July 5;406(10498):52–62. PMID: 40543520

29. Leboulleux S, Bournaud C, Chougnet CN, Lamartina L, Zerdoud S, Cao CD, Catargi B, Dygai I, Kelly A, Barge ML, Vera P, Rusu D, Schneegans O, Roux J, Raymond P, Benisvy D, Eberle MC, Bidault S, Nascimento C, Bastie D, Giraudet AL, Bardet S, Moullec NL, Roudaut N, Drui D, Godbert Y, Zalzali M, Drutel A, Morel O, Velayoudom FL, Ghuzlan AA, Schlumberger M, Buffet C, Borget I. Thyroidectomy without radioiodine in patients with low-risk thyroid cancer: 5 years of follow-up of the prospective randomised ESTIMABL2 trial. The Lancet Diabetes & Endocrinology. Elsevier; 2025 Jan 1;13(1):38–46. PMID: 39586309

30. Tweed CD, Quartagno M, Clements MN, Turner RM, Nunn AJ, Dunn DT, White IR, Copas AJ. Exploring different objectives in non-inferiority trials. BMJ. 2024 June 17;385:e078000. PMCID: PMC11181107

31. Althunian TA, de Boer A, Klungel OH, Insani WN, Groenwold RHH. Methods of defining the non-inferiority margin in randomized, double-blind controlled trials: a systematic review. Trials. 2017 Mar 7;18(1):107.

32. Chow SC, Shao J, Wang H, Lokhnygina Y. Sample Size Calculations in Clinical Research. 3rd ed. New York: Chapman and Hall/CRC; 2017.