

Outcomes of Graves' Disease Patients Following Antithyroid Drugs, Radioactive Iodine, or Thyroidectomy as the First-line Treatment

Xiaodong Liu, MPH,* Carlos K. H. Wong, PhD,†‡ Wendy W. L. Chan, MBBS,§ Eric H. M. Tang, BSc,‡
Yu Cho Woo, MBChB,¶ Cindy L. K. Lam, MD,‡ and Brian H. H. Lang, MS*✉

Background: The long-term outcomes of first-line choice among ATD, RAI, and thyroidectomy for GD patients remain unclear.

Objective: To compare the long-term morbidity, mortality, relapse, and costs of GD patients receiving first-line treatment.

Methods: A population-based retrospective cohort of GD patients initiating first-line treatment with ATD, RAI, or thyroidectomy as a first-line primary treatment between 2006 and 2018 from Hong Kong Hospital Authority was analyzed. Risks of all-cause mortality, CVD, AF, psychological disease, diabetes, and hypertension were estimated using Cox proportional hazards regression models. The 10-year healthcare costs, change of comorbidities, and risk of relapse were compared across treatments.

Results: Over a median follow-up of 90 months with 47,470 person-years, 6385 patients (ATD, 74.93%; RAI, 19.95%; thyroidectomy, 5.12%) who received first-line treatment for GD were analyzed. Compared with ATD group, patients who had undergone surgery had significantly lower risks of all-cause mortality [hazard ratio (HR) = 0.363, 95% confidence interval (CI) = 0.332–0.396], CVD (HR = 0.216, 95% CI = 0.195–0.239), AF (HR = 0.103, 95% CI = 0.085–0.124), psychological disease (HR = 0.279, 95% CI = 0.258–0.301), diabetes (HR = 0.341, 95% CI = 0.305–0.381), and

hypertension (HR = 0.673, 95% CI = 0.632–0.718). Meanwhile, RAI group was also associated with decreased risks of all-cause mortality (HR = 0.931, 95% CI = 0.882–0.982), CVD (HR = 0.784, 95% CI = 0.742–0.828), AF (HR = 0.622, 95% CI = 0.578–0.67), and psychological disease (HR = 0.895, 95% CI = 0.855–0.937). The relapse rate was 2.41% in surgery, 75.60% in ATD, and 19.53% in RAI group. The surgery group was observed with a significant lower Charlson Comorbidity Index score than the other 2 groups at the tenth-year follow-up. The mean 10-year cumulative healthcare costs in ATD, RAI, and surgery group was US\$23915, US\$24260, and US\$20202, respectively.

Conclusions: GD patients who received surgery as an initial treatment appeared to have lower chances of all-cause mortality, CVD, AF, psychological disease, diabetes, and hypertension in the long-term when compared to those treated with ATD or RAI. The surgery group had the lowest relapse and direct healthcare costs among the 3 treatment modalities. This long-term cohort study suggested surgery may have a larger role to play as an initial treatment for GD patients.

Keywords: antithyroid drugs, Graves' disease, population based cohort, radioactive iodine, thyroidectomy

(*Ann Surg* 2021;273:1197–1206)

From the *Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; †Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; ‡Department of Family Medicine and Primary Care, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; §Department of Clinical Oncology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; and ¶Department of Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.

✉blang@hku.hk

X.L. and C.K.H.W. contributed equally as first author.

Contribution statement: XL contributed to review the articles, analyzed the data, and wrote the manuscript. CKHW had the original idea for the study and contributed to the development of the study design. WWLC provided essential suggestions on the treatment of radioactive iodine from the clinician's point. EHMT provided technical support. BHHL provided critical feedback and revisions to the introduction, methodology, and conclusions of the manuscript and supervised the work. All authors contributed to the interpretation of analysis and provided comments to revise the manuscript.

Funding Sources: This study was funded by the Health and Medical Research Fund, Food and Health Bureau, Hong Kong SAR (Ref No. #17181051). No funding organization had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation of the manuscript.

The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsurgery.com).

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc.

ISSN: 0003-4932/21/27306-1197

DOI: 10.1097/SLA.00000000000004828

Graves' disease (GD) is a common thyroid disorder with a population prevalence of 1%–1.5%, occurring in approximately 3% of women and 0.5% of men in their lifetime.¹ It is the most frequent cause of hyperthyroidism.^{2,3} Treatment options for GD currently comprise antithyroid drugs (ATD), radioactive iodine (RAI), and surgery. Although each treatment has its own benefits and shortcomings, ATD treatment represents the predominant first-line treatment in Europe, Asia, and to some extent, the USA.^{4,5} Apart from the benefits of low cost and ease of administration, a continued ATD treatment of 12–18 months could enhance disease remission when compared to no intervention.⁶ Unlike ATD, RAI treatment utilizes ionizing radiation to cause cellular damage and deaths of thyroid follicular cells and decreases thyroid function and size of thyroid gland.⁷ Within 3–12 months after RAI treatment, the thyroid function among 50%–90% of patients is normalized.⁸ Surgery in the form of thyroidectomy is the least commonly-selected first-line treatment in GD patients. Previous surveys conducted in the USA and Europe reported that only around 2% of patients underwent surgery as a first-line therapy.^{4,9} Despite its benefits of rapid control of hyperthyroidism, lack of radiation exposure, and less chance of worsening coexisting Graves' ophthalmopathy, surgery requires hospitalization and bears permanent hypothyroidism, anesthetic, and surgical risks.¹⁰

However, it still remains unclear how the choice of first-line treatment may fully affect the patients' outcomes in the longer-term. It is known that GD patients with fair hyperthyroidism control are at increased risks of morbidity and mortality over time.¹¹ Some GD patients with poorly-controlled hyperthyroidism can develop heart-

related complications and diseases.¹² Recent studies reported that patients managed by either ATD or RAI were at increased risk of cardiovascular disease (CVD) and all-cause mortality when compared to those managed by surgery.^{13,14} A recent study suggested that surgery could modify long-term cardiac risks by improving co-existing hypertension and arrhythmias.¹⁵ Therefore, it is becoming evident that surgery may have a reduced risk of cardiac events in the long-term.^{13,16–20} Based on above evidence, we hypothesized that surgery as first-line treatment may lower the morbidity and all-cause mortality risks and this, in turn, lead to reduced overall healthcare costs in the long-term when compared to ATD or RAI.

The aim of our study was to compare the long-term outcomes of all-cause morbidity, CVD, AF, psychological disease, diabetes, and hypertension for GD patients receiving ATD, RAI, or surgery/thyroidectomy as first-line treatment. The 10-year direct healthcare costs, change of comorbidity profiles, and risk of relapse across different treatment modalities also were evaluated in the comparison.

METHODOLOGY

Study Population

This retrospective cohort study was approved by the local institutional review board. All data were retrieved from the territory-wide prospectively-coded database (the Hong Kong Hospital Authority Clinical Management System) which came into operation in 1998. This system is a computerized database linking up all 41 public hospitals and clinics and covers over 90% of all inpatient services in the territory. Given that some data were not fully validated until 2005, for the present study, data were retrieved from January 1st, 2006 to December 31st, 2018. All patients aged ≥ 18 years with a diagnosis of GD and received first-line treatment either in the form of ATD, RAI, or thyroidectomy were analyzed. The diagnosis of GD was identified using the International Statistical Classification of Diseases and Related Health Problems, Ninth Revision, Clinical Modification codes 242.00/242.01. Patients were excluded if any of the following criteria were met: had no hospital treatment records, lost to follow-up after baseline, or had a pregnancy within 12 months before the index date. Medical records extracted from the documented database consisted of clinical parameters, treatment modalities, laboratory tests, comorbidities, drug dispensation, and healthcare service utilization.

Our study focused mainly on 3 major treatment modalities, namely ATD, RAI, and thyroidectomy. The index date of subjects included was defined as the date when patients received first-line treatment. The follow-up period was ceased at the date of mortality, event occurrence, or censored at the final healthcare service utilization date. Among these dates, the earliest one was chosen to estimate each event outcome.

Assignment of Treatment Group

Given that it was recognized patients would take a brief period of ATD before definitive RAI or surgery, this period could vary up to 12 months depending on the institution and its workload. A 12-month or less period of ATD before RAI or thyroidectomy was permitted before the assignment of treatment group.²¹ In other words, patients were only assigned to the ATD group if they did not receive definitive treatment of RAI or thyroidectomy 12 months after ATD initiation. Patients who conducted RAI without receiving ATD, or those who received ATD for less than 12 months then changed to RAI therapy were assigned to the RAI group. Similarly, patients who performed thyroidectomy without receiving ATD, or those who received ATD for less than 12 months then changed to thyroidectomy were assigned to the surgery group.²²

Outcome Measures

Primary outcomes of this study were risks of CVD, AF, psychological disease, diabetes, hypertension, and all-cause mortality following first-line GD treatment. Diagnosis codes of International Statistical Classification of Diseases and Related Health Problems, Ninth Revision, Clinical Modification and International Classification of Primary Care, Version 2 were used to identify the events.

The risk of relapse after first-line treatment was estimated across groups until the end of follow-up. Association between dose of RAI and primary outcomes in the RAI therapy were also assessed. Additional outcomes included the 10-year direct cumulative healthcare costs incurred, and change of comorbidity profiles as estimated by Charlson Comorbidity Index (CCI) over time across treatment groups. CCI calculated by weighted scores assigned to 17 comorbidities such as myocardial infarction, heart failure, and other diseases, was applied to measure comorbid disease status as an indicator of disease burden.²³ Direct healthcare costs comprised the costs of clinic visits, hospitalization, surgical procedures, and relevant medications. All of the diagnosis and procedure codes used in the study are listed in Supplemental Table 3, <http://links.lww.com/SLA/C989>.

Statistical Analysis

Baseline covariates consisted of age and sex, clinical parameters, disease status, and previous medication treatment. Clinical parameters comprised thyroid function-related indicators of serum thyroid-stimulating hormone (TSH), blood pressure (systolic and diastolic), cholesterol ratio (total cholesterol divided by high-density lipoprotein), triglyceride, fasting blood sugar level, and estimated glomerular filtration rate (eGFR). All clinical parameter covariates were continuous except eGFR (< 60 , ≥ 60 mL/min/1.73 m²). Disease status included duration of GD, CCI, and history of cardiovascular morbidities like coronary heart disease, AF or heart failure, cerebrovascular disease, hypertension or diabetes on or before the initiation of first-line therapy. Previous medication treatment on or before the index date included drugs of beta-blocker, calcium-channel blocker, or corticosteroids.

Multiple imputations by chained equations were performed in adequately dealing with missing baseline covariate values.^{24,25} Rubin combination rules were applied to calculate the linear predictions with multiple imputed dataset.²⁶ Inverse probability of treatment weights was used to equilibrate the baseline variables of the patients in ATD, RAI, and thyroidectomy group, by estimating the propensity scores from multinomial logistic regression. A syntax created for marginal mean weighting was applied through hierarchical coding to conduct the inverse probability of treatment weights using the propensity scores.²⁷ The lowest and highest 1% propensity score weights were trimmed to obtain better balance between treatment groups. To estimate the equilibrium of variables among the 3 groups at baseline, propensity score weightings were repeated 5 times. Using imputed datasets, absolute standardized mean difference (ASMD) was estimated, the maximum value of which less than 0.2 indicating optimal balance. To minimize the influence caused by residual covariate imbalance, multivariable regression models were conducted repeatedly with an SMD value > 0.10 after propensity-score weighting and the models were double adjusted for variables.²⁸

The baseline characteristic data of subjects was summarized by using mean \pm standard error and frequencies (percentages) to present for continuous and categorical variables respectively after propensity score weighting. Univariate linear regression and binary logistic or multinomial logistic regression were performed to compare the differences of variables among treatment groups.

Computing the number of patient cases in events divided by person-years during follow-up, the incidence rate of each study outcome was calculated in different groups. Based on the assumption that the incident cases followed the Poisson distribution, the 95% confidence interval (CI) was estimated for the incidence rate. Kaplan-Meier curves and Cox hazard proportional regression models adjusted for the baseline covariates were applied to compare the risks of event outcomes between treatments. To estimate the treatment effects for event outcomes, hazard ratios (HRs) combined with 95% CI were computed. Equality in survival functions between treatment groups was compared using the log-rank test.

The risk of GD relapse was compared between treatment modalities. For patients in the first-line ATD group, disease relapse was defined as a patient requiring a continuing maintenance dose of ATD for 18 months or longer or receiving RAI or thyroidectomy following a complete 18 months or longer course of ATD therapy. For patients in the first-line RAI group, disease relapse was defined as a patient who continued to require or restarted ATD therapy within 12 months after the administration of the first RAI dose. If a patient received 1 or more subsequent RAI dose or needed thyroidectomy afterward, it was also considered a disease relapse. For patients in the first-line surgery group, relapse was defined as a patient who continued to require or restarted ATD within 12 months of surgery or received RAI or additional thyroid surgery subsequently. Among patients receiving RAI as first-line treatment, association between dose of RAI and primary outcomes was also estimated. Odds ratio (OR) and 95% CI were computed by multivariable logistic regression models to compare the risk of GD relapse across treatment groups and assess the relationship between RAI dose with primary outcomes.

The change of comorbidities over time and 10-year healthcare costs across groups were compared. Change of comorbidities was presented by the proportion of patients in respective categories of CCI. Direct cumulative healthcare costs consisted of 2 parts: health service cost calculated by the unit cost of service and the number of services used, medication cost calculated by the unit cost of drug and the number of drug dispensation extracted from the dataset. The unit cost of healthcare services and drugs are listed in Supplemental Table 4, <http://links.lww.com/SLA/C989>. Mean CCI and direct cumulative healthcare costs from baseline to 10-year follow-up were estimated for each treatment group. *P*-values on the effect of overall treatments on CCI and healthcare costs at baseline, 1-year to 10-year were reported in figures, which were calculated by the multivariable linear regression.

Sensitivity Analysis

The primary analysis of our study where treatments were assigned and outcome data was available for all subjects was conducted with "Intention-to-treat." Sensitivity analyses comprising complete-case, "as-treated," and competing risk analysis, and subgroup analyses were performed for the risks of primary outcomes. The complete-case analysis only included patients with no missing values of clinical data. "As-treated" analysis was defined as the estimation by the censoring time of first-line therapy or switching to another treatment modality. Competing risk analysis was performed to estimate the HRs of study outcomes of interest in the presence of competing risk event of all-cause mortality.²⁹

Subgroup analyses were performed by separating subjects into different subgroups according to variables, such as age and CCI at baseline, sex, and the TSH values at 12–18 months after treatments. The TSH values were divided into 3 levels: low level (<0.5 mIU/L), normal level (0.5–5 mIU/L), and high level (>5 mIU/L). *E*-value for sensitivity analysis was conducted to estimate the specific treatment-outcome association with the calculation of strength between

unmeasured confounding factors and treatment, and between confounding factors and outcome.³⁰ For the relapse rate, sensitivity analyses at censoring time of 2-year, 5-year, and 10-year follow-up were conducted to test the difference of results.

All statistical analyses were conducted using STATA version 16.0, (Stata Corp LP, College Station, TX). A 2-tailed *P* value < 0.05 was considered statistically significant.

RESULTS

The selection of subjects for this study is illustrated in the flowchart of Figure 1. Among 7223 patients with GD identified at Hospital Authority between 2006 and 2018, patients who did not have GD treatment records (*n* = 522), lost to follow-up after baseline (*n* = 68), under 18 years old (*n* = 147), or had been pregnant within 12 months before the index date (*n* = 101) were excluded. Of 6385 patients eligible for inclusion in the current analysis, 4784 (74.93%) patients received ATD, 1274 (19.95%) treated with RAI, and 327 (5.12%) patients underwent complete or partial thyroidectomy as first-line treatment.

Among patients with surgery, 59 (18.04%) conducted partial thyroidectomy and 268 (81.96%) performed complete thyroidectomy. Notably, laryngeal nerve paralysis occurred in 9.17%, hypocalcemia in 30.58%, hypoparathyroidism in 2.75%, and hematoma in 0.92% of patients in surgery group, respectively. The incidence of hypothyroidism was 7.58% for ATD, 40.85% for RAI, and 96.67% for thyroidectomy treatment at 12 months of follow-up. Over the 2 years, 2403 (50.23%) in the ATD group and 1003 (78.73%) in the RAI group achieved remission. Of those who did not achieve remission, 100 (4.2%) in the ATD group and 79 (28%) in the RAI group underwent surgery.

Baseline demographic characteristics of GD patients are presented in Supplemental Table 1, <http://links.lww.com/SLA/C989>. All covariates were balanced after multiple imputations and propensity score weighting. The mean age of all participants was 45.64 ± 0.52 years old, and 72.63% of patients were women. A total of 99.07% of all patients had a GD duration of no more than 18 months before initiating first-line treatment. Details of baseline characteristics before weighting are listed in Supplemental Table 5, <http://links.lww.com/SLA/C989>.

The cumulative incidence and incidence rate of all-cause mortality, CVD, AF, psychological disease, diabetes, and hypertension by first-line treatment groups are presented in Supplemental Table 2, <http://links.lww.com/SLA/C989>. Over a median follow-up of 90 months with 47,470 person-years, the incidence rates of all-cause mortality, CVD, AF, psychological disease, diabetes, and hypertension were the lowest in the surgery group (5.53, 2.79, 0.80, 5.24, 2.50, and 10.61 per 1000 person-years respectively). Compared with ATD group, RAI treatment was observed to have lower incidence rates of all-cause mortality (10.65 vs 12.47 per 1000 person-years), CVD (10.02 vs 13.81 per 1000 person-years), AF (4.95 vs 9.14 per 1000 person-years), and psychological disease (15.45 vs 17.54 per 1000 person-years), but higher incidence rates of diabetes (8.61 vs 8.29 per 1000 person-years) and hypertension (21.60 vs 17.96 per 1000 person-years). Among these outcomes, all-cause mortality, CVD, and hypertension accounted for the top 3 rank positions, which was consistent across different groups.

The risks of primary outcomes estimated by Cox regression models adjusted by baseline covariates are shown in Table 1. Compared with ATD group, thyroidectomy was associated with significantly lower risks of all-cause mortality (HR = 0.363, 95% CI = 0.332–0.396), CVD (HR = 0.216, 95% CI = 0.195–0.239), AF (HR = 0.103, 95% CI = 0.085–0.124), psychological disease (HR = 0.279, 95% CI = 0.258–0.301), diabetes (HR = 0.341, 95% CI =

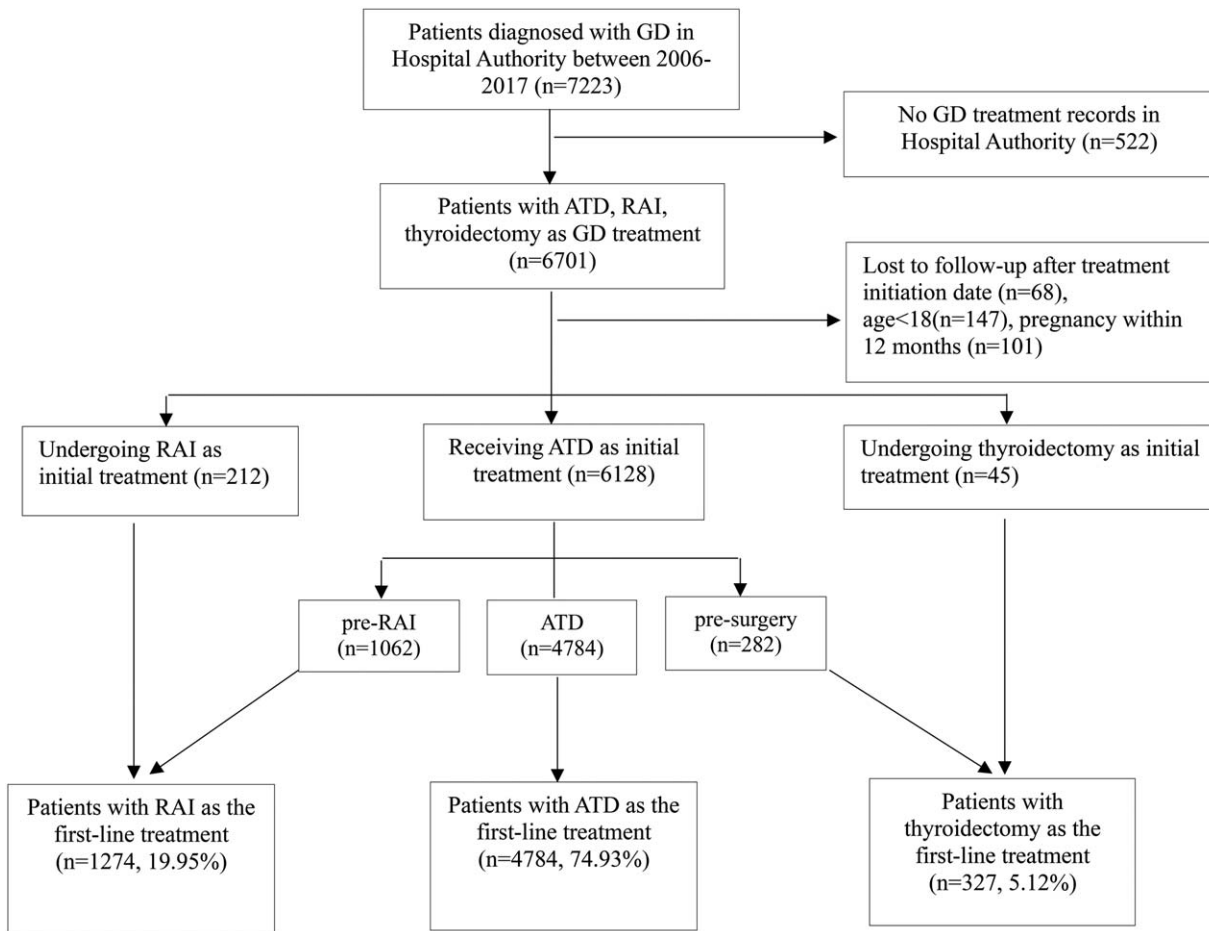


FIGURE 1. Flowchart of Graves’ disease patients with ATD, RAI, and surgery as first-line treatment. ATD indicates antithyroid drug; RAI, radioactive iodine.

0.305–0.381), and hypertension (HR = 0.673, 95% CI = 0.632–0.718). Meanwhile, the RAI group was found to have reduced risks of all-cause mortality (HR = 0.931, 95% CI = 0.882–0.982), CVD (HR = 0.784, 95% CI = 0.742–0.828), AF (HR = 0.622, 95% CI = 0.578–0.67), and psychological disease (HR = 0.895, 95% CI = 0.855–0.937), but higher risks of diabetes (HR = 1.081, 95% CI = 1.014–1.152) and hypertension (HR = 1.255, 95% CI = 1.203–1.31) than patients on ATD. The comparison between surgery and RAI groups also indicated significantly lower risks of all-cause mortality (HR = 0.446, 95% CI = 0.408–0.487), CVD (HR = 0.296, 95% CI = 0.267–0.328), AF (HR = 0.256, 95% CI = 0.212–0.308), psychological disease (HR = 0.303, 95% CI = 0.28–0.327), diabetes (HR = 0.269, 95% CI = 0.242–0.3), and hypertension (HR = 0.555, 95% CI = 0.523–0.589) for those receiving thyroidectomy as first-line treatment. The majority of results obtained from the sensitivity analyses of complete cases, “as-treated,” and competing risk were consistent with those of primary analysis (Supplemental Table 6, <http://links.lww.com/SLA/C989>). In the “as-treated” analysis, however, the RAI group was associated with lower risks of diabetes (HR = 0.520, 95% CI = 0.480–0.564) and hypertension (HR = 0.602, 95% CI = 0.569–0.638) than that of ATD.

The risk of relapse by the end of the follow-up was calculated across treatment groups (Table 2). Patients who had undergone thyroidectomy were observed to have the lowest relapse rate

(2.41% vs 19.53% with RAI or 75.60% with ATD). Furthermore, the surgery (OR = 0.010, 95% CI 0.009–0.011) and RAI (OR = 0.084, 95% CI = 0.081–0.087) groups were associated with significantly reduced risk of relapse compared with the ATD group at the censoring of follow-up. Similar results were found in the sensitivity analyses at different endpoints of 2-year, 5-year, and 10-year follow-up. The relapse rate of patients after partial thyroidectomy was 4.16% compared to the rate of 2.21% after complete thyroidectomy.

The associations between dose of RAI and developing event outcomes for patients in the first-line RAI group are presented in Table 3. Compared with patients who received 1 dose of RAI, the increased odds of all-cause mortality (OR = 2.790, 95% CI = 1.761–4.422) and hypertension (OR = 1.793, 95% CI = 1.125–2.858) were observed in those with 2 doses of radioactive activity. No significant differences were found in CVD, AF, psychological disease, and diabetes.

Figure 2 depicts the Kaplan-Meier survival curves of all-cause mortality, CVD, AF, psychological disease, diabetes, and hypertension over a 10-year period for each treatment group. Distributions of all outcomes were significantly different among patients treated with ATD, RAI, or thyroidectomy. Compared with patients receiving ATD or RAI, those undergoing thyroidectomy were observed to have the lowest risk of all outcomes. The Kaplan-Meier survival curves of coronary heart disease, heart failure, and stroke suggesting lower

TABLE 1. Hazard Ratios of All-cause Mortality, Cardiovascular Disease, Atrial Fibrillation, Psychological Disease, Diabetes, and Hypertension by Treatment Groups

Event	Surgery Versus ATD			Surgery Versus RAI			RAI Versus ATD		
	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value
Primary analysis – “Intention-to-treat”									
All-cause mortality	0.363	(0.332, 0.396)	<0.001*	0.446	(0.408, 0.487)	<0.001*	0.931	(0.882, 0.982)	0.008*
CVD	0.216	(0.195, 0.239)	<0.001*	0.296	(0.267, 0.328)	<0.001*	0.784	(0.742, 0.828)	<0.001*
AF	0.103	(0.085, 0.124)	<0.001*	0.256	(0.212, 0.308)	<0.001*	0.622	(0.578, 0.67)	<0.001*
Psychological disease	0.279	(0.258, 0.301)	<0.001*	0.303	(0.280, 0.327)	<0.001*	0.895	(0.855, 0.937)	<0.001*
Diabetes	0.341	(0.305, 0.381)	<0.001*	0.269	(0.242, 0.3)	<0.001*	1.081	(1.014, 1.152)	0.017*
Hypertension	0.673	(0.632, 0.718)	<0.001*	0.555	(0.523, 0.589)	<0.001*	1.255	(1.203, 1.31)	<0.001*
Sensitivity analysis – As-treated analysis									
All-cause mortality	0.368	(0.325, 0.417)	<0.001*	0.537	(0.489, 0.589)	<0.001*	0.911	(0.841, 0.987)	0.023*
CVD	0.133	(0.119, 0.149)	<0.001*	0.323	(0.290, 0.360)	<0.001*	0.408	(0.381, 0.437)	<0.001*
AF	0.084	(0.069, 0.102)	<0.001*	0.318	(0.261, 0.389)	<0.001*	0.365	(0.334, 0.399)	<0.001*
Psychological disease	0.167	(0.153, 0.183)	<0.001*	0.288	(0.266, 0.313)	<0.001*	0.489	(0.46, 0.52)	<0.001*
Diabetes	0.183	(0.162, 0.207)	<0.001*	0.240	(0.215, 0.269)	<0.001*	0.520	(0.480, 0.564)	<0.001*
Hypertension	0.424	(0.392, 0.458)	<0.001*	0.634	(0.594, 0.676)	<0.001*	0.602	(0.569, 0.638)	<0.001*

*Significant at 0.05 level by multivariable Cox proportional hazard regression.

AF indicates atrial fibrillation; ATD, antithyroid drugs; CI, confidence interval; CVD, cardiovascular diseases; HR, hazard ratio; RAI, radioactive iodine.

risks in surgery group compared to the ATD and RAI group are shown in the Supplemental Figure 1, <http://links.lww.com/SLA/C989>.

Across treatment groups, the increasing proportions of high CCI value categories and the growing trend of CCI scores over the follow-up period were illustrated in Figure 3. Compared to the ATD and RAI group, the CCI score of the surgery group appeared to be slightly higher in the first several years but decreased more after propensity score weighting. A significantly lower CCI score was identified in patients undergoing thyroidectomy at the tenth-year follow-up. The changes of TSH, FT4, serum creatinine, and eGFR in

different groups are shown in Supplemental Figure 2, <http://links.lww.com/SLA/C989>.

The annual and cumulative direct healthcare costs comprising healthcare service utilization and medication use over a 10-year follow-up are displayed in Figure 4. Following the highest annual expenses incurred in the first year, a decreasing trend was observed over time across all treatment groups. The 10-year cumulative healthcare cost of the surgery group (US\$20202) was relatively lower than that of ATD (US\$23915) and RAI (US\$24260). Notably, the cumulative cost of patients in the surgery group became similar to those in the ATD and RAI groups after the fifth year. The frequency

TABLE 2. Disease Relapse by Treatment Groups of ATD, RAI, and Surgery

Groups	Rate	Median Time to Relapse (month)	OR	95% CI	P value
Total	35.54%	49			
ATD	75.60%	21		1 (Reference)	
RAI	19.53%	80	0.084	(0.081, 0.087)	<0.001*
Surgery	2.41%	82	0.010	(0.009, 0.011)	<0.001*
Sensitivity analysis-censoring at the 2 yr of follow-up					
Total	23.69%	24			
ATD	54.78%	21		1 (Reference)	
RAI	8.16%	24	0.076	(0.073, 0.080)	<0.001*
Surgery	2.13%	24	0.022	(0.020, 0.024)	<0.001*
Sensitivity analysis-censoring at the 5 yr of follow-up					
Total	32.88%	49			
ATD	72.25%	21		1 (Reference)	
RAI	15.80%	60	0.070	(0.072, 0.078)	<0.001*
Surgery	2.13%	60	0.010	(0.009, 0.011)	<0.001*
Sensitivity analysis-censoring at the 10 yr of follow-up					
Total	35.44%	49			
ATD	75.50%	21		1 (Reference)	
RAI	19.36%	80	0.083	(0.080, 0.086)	<0.001*
Surgery	2.41%	82	0.010	(0.009, 0.011)	<0.001*

*Significant at 0.05 level by multivariable logistic regression.

(1) For patients in the ATD group, relapse was defined as receiving continuous ATD after a course of 18-mo first-line ATD treatment, or conducting subsequent treatment of RAI or thyroidectomy.

(2) For patients in the RAI group, relapse was defined as the conduction of the second RAI or subsequent treatment of thyroidectomy or required a maintenance dose of ATD after the first RAI treatment.

(3) For patients in the surgery group, relapse was defined as the second thyroidectomy, or subsequent treatment of ATD or RAI.

Patients who received a less than 12 mo period of ATD then changed to RAI or thyroidectomy would be defined as pre-RAI or pre-thyroidectomy.

ATD indicates antithyroid drugs; CI, confidence interval; OR, odds ratio; RAI, radioactive iodine.

TABLE 3. Comparison of Radiation Dose at Subsequent Risks of Mortality, Cardiovascular Disease, Atrial Fibrillation, Psychological Disease, Diabetes, and Hypertension

Event	OR [†]	95% CI	P value
All-cause mortality	5.181	(1.277, 21.023)	0.021*
Cardiovascular disease	0.473	(0.049, 4.554)	0.517
Atrial fibrillation	NA [‡]	NA	NA
Psychological disease	0.594	(0.068, 5.167)	0.637
Diabetes	4.376	(0.748, 25.619)	0.102
Hypertension	1.971	(0.302, 12.841)	0.478
Sensitivity analysis before adjustment [§]			
All-cause mortality	3.706	(2.175, 6.317)	<0.001*
Cardiovascular disease	0.926	(0.530, 1.619)	0.788
Atrial fibrillation	1.637	(0.570, 4.702)	0.360
Psychological disease	0.667	(0.264, 1.687)	0.393
Diabetes	1.524	(0.677, 3.434)	0.309
Hypertension	2.542	(1.479, 4.367)	0.001*

*Significant at 0.05 level by multivariable logistic regression.

[†]Comparison between patients who received 2 doses of RAI and those with 1 dose of RAI (reference).

[‡]NA: atrial fibrillation occurred in patients who had one RAI dose.

[§]Adjusted for variables of age, sex, thyroid-stimulating hormone, free thyroxine, systolic blood pressure, diastolic blood pressure, total cholesterol/HDL cholesterol ratio, triglyceride, fasting glucose, estimated glomerular filtration rate, methimazole, propylthiouracil, carbimazole, duration, and Charlson Comorbidity Index.

CI indicates confidence interval; NA, not available; OR, odds ratio; RAI, radioactive iodine.

of health utilization by treatment groups is shown in Supplemental Figure 3, <http://links.lww.com/SLA/C989>, also indicating the highest healthcare service used in the first year and a decreasing trend in the following years.

In general, the majority of results in subgroup analyses stratified by age, sex, CCI, and TSH level were in line with those of primary analysis (Supplemental Table 7, <http://links.lww.com/SLA/C989>). A few discrepancies existed in the comparison of results in certain patient subgroups. For patients treated with RAI, a higher risk of all-cause mortality was observed in those aged <60 years (HR = 1.125, 95% CI = 1.028, 1.23), who were female (HR = 1.121, 95% CI = 1.051, 1.196), or with CCI <4 (HR = 1.106, 95% CI = 1.018, 1.203), but a lower risk of diabetes was identified in those aged >=60 years (HR = 0.856, 95% CI = 0.747, 0.981) or who were female (HR = 0.911, 95% CI = 0.84–0.988), when compared to the ATD group. Among patients who had normal TSH level at 12–18 months after treatment, patients treated with RAI had higher risks of all-cause mortality (HR = 1.159, 95% CI = 1.044–1.287), AF (HR = 1.293, 95% CI = 1.094–1.527), and psychological disease (HR = 1.124, 95% CI = 1.022–1.235); but lower risks of diabetes (HR = 0.825, 95% CI = 0.723–0.942) and hypertension (HR = 0.854, 95% CI = 0.782–0.934) than those on ATD. Meanwhile, among patients with low TSH level, those undergoing surgery had a higher risk of diabetes (HR = 1.311, 95% CI = 1.086–1.582) compared with those taking ATD. Estimates of *E*-values were larger than all of the HRs for all-cause mortality, CVD, AF, psychological disease, diabetes, and hypertension, suggesting that there was a minimal chance of unknown confounders having greater effects on the study outcomes than the baseline variables on the outcomes (Supplemental Table 8, <http://links.lww.com/SLA/C989>).

DISCUSSION

Consistent with our initial hypothesis, relative to ATD and RAI therapy, patients undergoing initial surgery had significantly lower risks of all-cause mortality, CVD, AF, psychological disease,

diabetes, and hypertension in the long-term. Besides, as a result of the lower relapse rate, the 10-year healthcare cost in the surgery group was among the lowest when compared to ATD or RAI as a first-line treatment.

Giesecke et al reported that patients undergoing thyroidectomy treatment for hyperthyroidism was associated with reduced risks of all-cause and CVD mortality when compared to RAI therapy.¹³ Another study also found that patients after thyroidectomy had a lower CVD risk than that of those treated with RAI.³¹ These findings were largely consistent with our results reporting that surgery resulted in lower mortality and CVD risks than RAI therapy. Meanwhile, compared to ATD, significant clinical improvement was observed in the performance of hypertension, tachyarrhythmias including AF, and heart failure after thyroidectomy therapy among GD patients with CVD comorbidity.¹⁴ This further echoed our finding that surgery had lower CVD risk than ATD treatment in the long-term.

Our study reported that patients with RAI treatment had decreased mortality and CVD risks compared to those treated with ATD. This was in concordance to a cohort study where RAI treatment was associated with lower risks of mortality and major cardiovascular events including AF compared to ATD therapy.²⁰ However, GD patients might be more prone to hypothyroidism after RAI treatment and that may result in worsened insulin resistance and diabetes. Such reasons may contribute to explain the higher risk of diabetes in RAI than ATD group in the main analysis.³² The studies focusing on psychological diseases caused by GD treatments are lacking. From the research published, psychological disorder due to impaired vision and change of appearance is common in GD patients with ophthalmopathy.³³ RAI treatment was reported to increase the risk of developing or worsening ophthalmopathy that may lead to a higher occurrence of psychological disease. The result in our study where a lower risk of psychological disorder occurred in RAI compared to ATD group supplemented new evidence for the impact comparison between GD treatments on psychological diseases. Research comparing the influence between ATD and RAI treatments on the event of hypertension is limited, in consequence, more individual studies need to be explored by researchers.

Among the 3 treatment modalities, the surgery group incurred the least healthcare expense over the 10-year period. This was despite the fact that surgery resulted in a higher annual cost in the first year. This long-term cost-saving might have been a result of the lower relapse rate which led to a lower chance of continuous monitoring and treatment including definitive RAI and surgery later. Regarding cost-effectiveness as an important factor for consideration in the choice of GD therapy, our results favored surgery with lower long-term healthcare costs. This finding was in line with a lifetime analysis where total thyroidectomy was more cost-effective than RAI treatment or lifelong ATD.³⁴ In our study, the CCI value of the surgery group was observed to decrease more, especially a lower score was found compared to RAI and ATD therapy when a significant difference was identified at the tenth-year follow-up. The risks of primary outcomes in the surgery group were lower than those of the other 2 groups, contributing to explain difference in comorbidities and healthcare costs.

The lower recurrence associated with RAI treatment than that of ATD in our study was in line with a meta-analysis where patients after RAI were found to have a reduced relapse rate and increased cure rate compared to those treated with ATD therapy.³⁵ Compared to those treated with subtotal thyroidectomy, patients receiving total or near-total thyroidectomy are more likely to have no recurrence after the operation.³⁶ The relapse rates of 75.60% in ATD, 19.53% in RAI, and 2.41% in the surgery group of our study were similar to those of 52.7%, 15%, and 10% as previously reported by a meta-analysis.³⁷

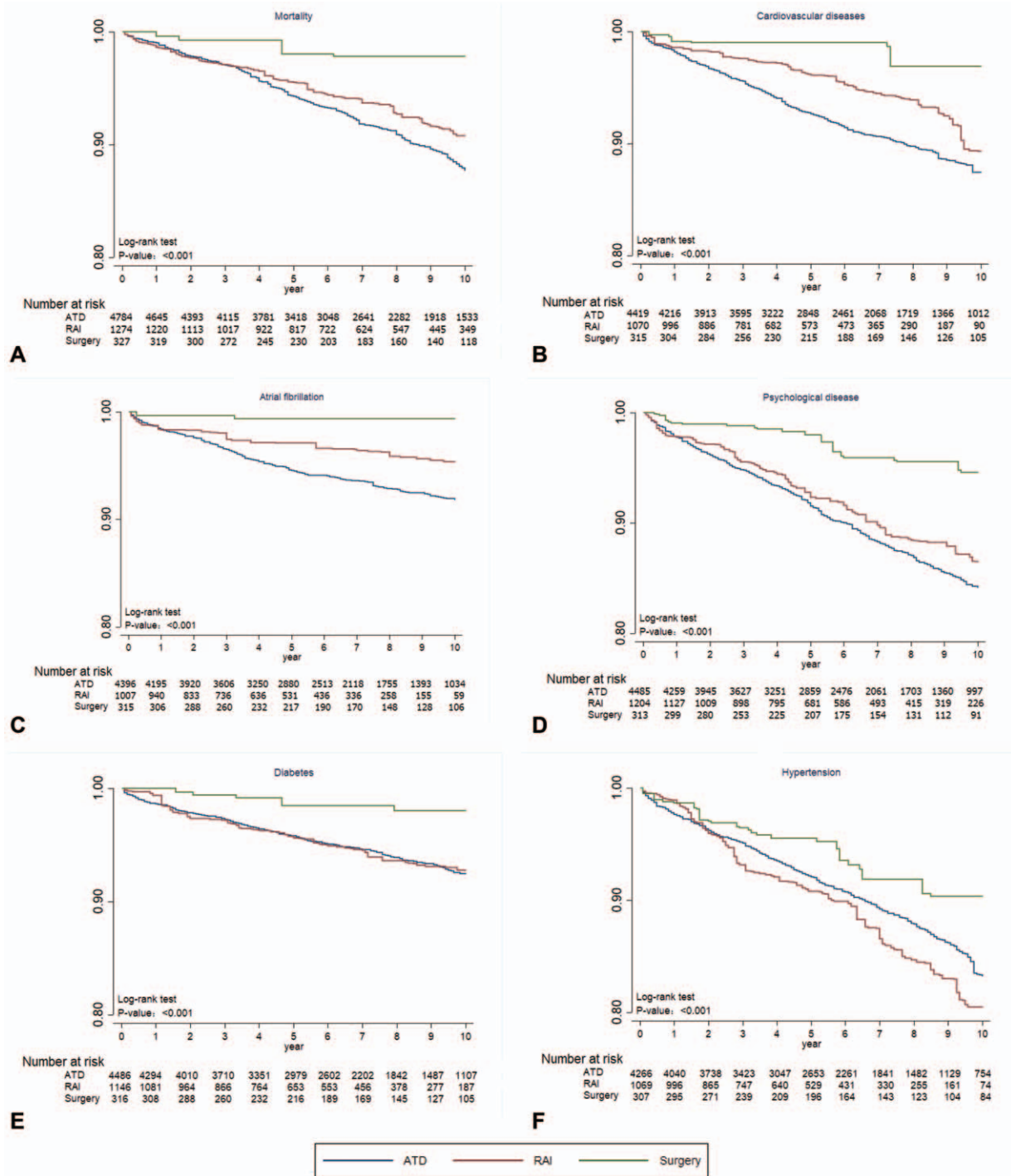
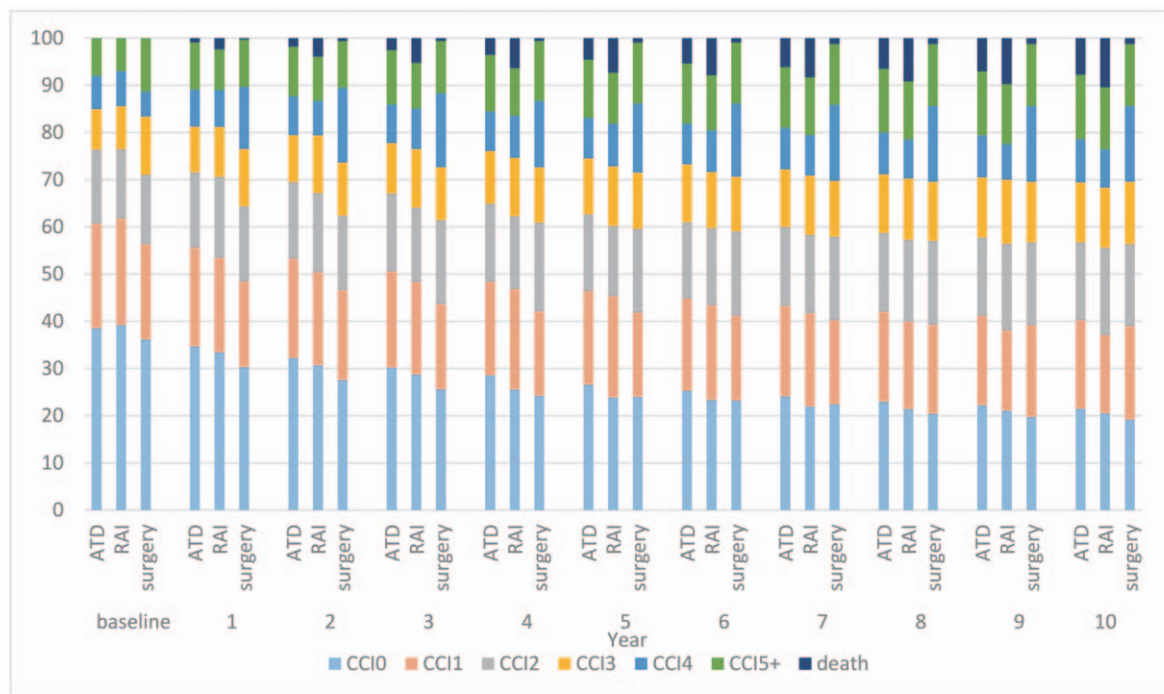


FIGURE 2. Kaplan-Meier curves of mortality (A), cardiovascular disease (B), atrial fibrillation (C), psychological disease (D), diabetes (E), and hypertension (F) for patients with ATD, RAI, and surgery. ATD indicates antithyroid drug; RAI, radioactive iodine.

One previous study surveyed the ATD regimens after RAI in different hospitals and found that a 2–6 months period of ATD therapy was combined usually for patients after 131-I therapy.³⁸ The condition that patients received ATD after first-line treatment of RAI should be taken into consideration. Besides, nearly 82% of patients performed

complete thyroidectomy which might lead to a lower relapse rate in our study.

Administered dose of ionizing radiation is one of the important considerations in the choice of first-line treatment for GD, in view of their potential impacts on various clinical outcomes. Our



ATD	1.47	1.69	1.80	1.92	2.03	2.13	2.21	2.28	2.34	2.39	2.44
RAI	1.44	1.70	1.82	1.92	2.01	2.10	2.17	2.23	2.29	2.34	2.38
Surgery	1.64	1.91	2.01	2.09	2.14	2.20	2.23	2.27	2.31	2.32	2.33
P-value	0.936	0.231	0.379	0.521	0.525	0.600	0.382	0.254	0.184	0.093	0.040*

FIGURE 3. Ten-years change in Charlson Comorbidity Index for patients with ATD, RAI, and surgery. ATD indicates antithyroid drug; RAI, radioactive iodine. *Significant at 0.05 level by multivariable linear regression.

study estimated that patients who had been treated with 2 doses of RAI had a higher risk of all-cause mortality, this aligned with 2 cohort studies where a positive association between RAI dose and mortality was observed.^{39,40} Besides, factors such as age, sex, CCI, and TSH level were found to influence the impacts of GD treatments on clinical outcomes. Most of our results comparing RAI and ATD treatments in the primary analysis were different from those calculated among patients with normal TSH concentration at 12–18 months in subgroup analysis. Differences in such outcomes might be connected with that biochemical euthyroidism with normal TSH at an early time was warranted particularly for patients treated with ATD as first-line treatment. The result of all-cause mortality estimated among RAI and ATD patients with low CCI values in our study was consistent with that of a previous cohort study where reduced mortality occurred in patients treated with the ATD of thionamide compared to 131-I therapy.⁴¹ The result in the previous study was observed in patients without significant comorbidities, indicating that the risk of mortality might favor ATD therapy over RAI when patients had few comorbidities before treatment initiation.

There are a number of strengths in our study. This study assessed the outcomes of all-cause mortality, CVD, AF, psychological disease, diabetes, and hypertension across 3 treatment modalities as first-line treatments for GD patients during a long follow-up period. Besides, the risk of relapse, change of comorbidities, and direct healthcare costs were also estimated and compared across treatment groups, so as to provide more comprehensive information on the choice of first-line GD therapies. Baseline covariates including demographic characteristics and comorbidities were balanced

and adjusted in the estimation of event outcomes, improving the precision of results estimated. Nevertheless, some limitations of this study need to be pointed out. Firstly, attributing to data availability of the data source, factors such as free triiodothyronine, total triiodothyronine, total thyroxine, and body mass index, were insufficient for the current analysis. Thus, the impact of these factors could not be accounted for. Secondly, smoking recognized as one of the important confounders in mortality, CVD and GD⁴²; was not available in the database, hence could not be adjusted for outcomes analyzed. Such bias might result in higher risks of event outcomes than reality, however, the E-values estimated for all outcomes suggested that it was unlikely to have substantial influence caused by unobserved confounders on our observed results compared to the covariates included in this study. Lastly, the direct healthcare costs were calculated based on the local healthcare service system and payment structure, which was limited geographically to be applied to places having similar healthcare. However, our study presented and compared costs among different treatments over a long-term period to provide evidence for practitioners to assess their own performance.

CONCLUSIONS

Our data showed that thyroidectomy was associated with lower long-term risks than ATD and RAI in all-cause mortality, CVD, AF, psychological disease, diabetes, and hypertension. Similarly, RAI treatment was associated with reduced risks of all-cause mortality, CVD, AF, and psychological disease than ATD therapy. The surgery group also incurred the lowest 10-year accumulative

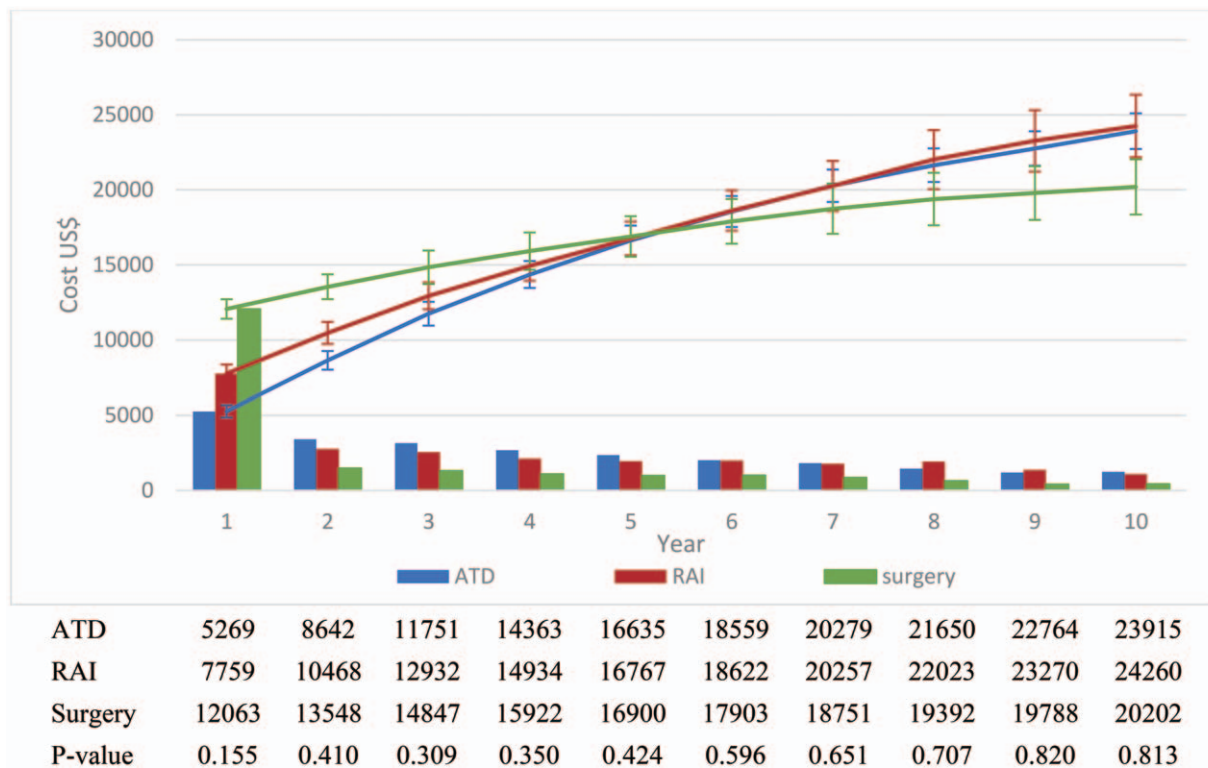


FIGURE 4. Annual and cumulative direct medical costs by groups of patients with ATD, RAI, and surgery. ATD indicates antithyroid drug; RAI, radioactive iodine. *Significant at 0.05 level by multivariable linear regression.

healthcare cost due to the lower relapse rate. There appeared to be an association between number of RAIs administered and long-term mortality risk. This study lends support to an increased role of surgery as a first-line treatment in GD.

ACKNOWLEDGMENT

The authors thank the Central Panel on Administrative Assessment of External Data Requests, Hong Kong Hospital Authority Head Office, for the provision of Hospital Authority data.

REFERENCES

- Nystrom HF, Jansson S, Berg G. Incidence rate and clinical features of hyperthyroidism in a long-term iodine sufficient area of Sweden (Gothenburg) 2003-2005. *Clin Endocrinol*. 2013;78:768-776.
- Franklyn JA, Boelaert K. Thyrotoxicosis. *Lancet*. 2012;379:1155-1166.
- Taylor PN, Albrecht D, Scholz A, et al. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol*. 2018;14:301-316.
- Bartalena L, Burch HB, Burman KD, et al. A 2013 European survey of clinical practice patterns in the management of Graves' disease. *Clin Endocrinol (Oxf)*. 2016;84:115-120.
- Smith TJ, Hegedus L. Graves' disease. *N Engl J Med*. 2016;375:1552-1565.
- Abraham P, Avenell A, McGeoch SC, et al. Antithyroid drug regimen for treating Graves' hyperthyroidism. *Cochrane Database Syst Rev*. 2010;1:CD003420.
- Mumtaz M, Lin LS, Hui KC, et al. Radioiodine I-131 for the therapy of Graves' disease. *Malays J Med Sci*. 2009;16:25-33.
- Bonnema SJ, Hegedus L. Radioiodine therapy in benign thyroid diseases: effects, side effects, and factors affecting therapeutic outcome. *Endocr Rev*. 2012;33:920-980.
- Burch HB, Burman KD, Cooper DS. A 2011 survey of clinical practice patterns in the management of Graves' disease. *J Clin Endocrinol Metab*. 2012;97:4549-4558.
- Bartalena L, Chiovato L, Vitti P. Management of hyperthyroidism due to Graves' disease: frequently asked questions and answers (if any). *J Endocrinol Invest*. 2016;39:1105-1114.
- Nyirenda MJ, Clark DN, Finlayson AR, et al. Thyroid disease and increased cardiovascular risk. *Thyroid*. 2005;15:718-724.
- Osman F, Gammage MD, Franklyn JA. Hyperthyroidism and cardiovascular morbidity and mortality. *Thyroid*. 2002;12:483-487.
- Giesecke P, Frykman V, Wallin G, et al. All-cause and cardiovascular mortality risk after surgery versus radioiodine treatment for hyperthyroidism. *Br J Surg*. 2018;105:279-286.
- Elnahla A, Attia AS, Khadra HS, et al. Impact of surgery versus medical management on cardiovascular manifestations in Graves disease. *Surgery*. 2021;169:82-86.
- Gauthier JM, Mohamed HE, Noureldine SI, et al. Impact of thyroidectomy on cardiac manifestations of Graves' disease. *Laryngoscope*. 2016;126:1256-1259.
- Sugrue D, McEvoy M, Feely J, et al. Hyperthyroidism in the land of Graves: results of treatment by surgery, radio-iodine and carbimazole in 837 cases. *Q J Med*. 1980;49:51-61.
- Ryodi E, Metso S, Jaatinen P, et al. Cancer incidence and mortality in patients treated either with RAI or thyroidectomy for hyperthyroidism. *J Clin Endocrinol Metab*. 2015;100:3710-3717.
- Sundaresh V, Brito JP, Thapa P, et al. Comparative effectiveness of treatment choices for graves' hyperthyroidism: a historical cohort study. *Thyroid*. 2017;27:497-505.
- Wu VT, Lorenzen AW, Beck AC, et al. Comparative analysis of radioactive iodine versus thyroidectomy for definitive treatment of Graves disease. *Surgery*. 2017;161:147-154.
- Okosieme OE, Taylor PN, Evans C, et al. Primary therapy of Graves' disease and cardiovascular morbidity and mortality: a linked-record cohort study. *Lancet Diabetes Endocrinol*. 2019;7:278-287.
- Rosato L, De Crea C, Bellantone R, et al. Diagnostic, therapeutic and healthcare management protocol in thyroid surgery: a position statement of the Italian Association of Endocrine Surgery Units (U.E.C. CLUB). *J Endocrinol Invest*. 2016;39:939-953.

22. Brito JP. Patterns of use, efficacy, and safety of treatment options for patients with Graves' disease: a nationwide population-based study (vol 30, pg 357, 2020). *Thyroid*. 2020;30:938–1938.
23. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992;45:613–619.
24. Raghunathan TE, Lepkowski JM, Hoewyk JV, et al. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*. 2001;27:85–95.
25. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007;16:219–242.
26. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: Wiley-Interscience; 2004.
27. Linden A. MMWS: Stata module to perform marginal mean weighting through stratification. Chestnut Hill, MA, Statistical Software Components, 2014.
28. Nguyen TL, Collins GS, Spence J, et al. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. *BMC Med Res Methodol*. 2017;17:78.
29. Satagopan JM, Ben-Porat L, Berwick M, et al. A note on competing risks in survival data analysis. *Brit J Cancer*. 2004;91:1229–1235.
30. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med*. 2017;167:268–274.
31. Ryodi E, Metso S, Huhtala H, et al. Cardiovascular morbidity and mortality after treatment of hyperthyroidism with either radioactive iodine or thyroidectomy. *Thyroid*. 2018;28:1111–1120.
32. Hage M, Zantout MS, Azar ST. Thyroid disorders and diabetes mellitus. *J Thyroid Res*. 2011;2011:439463.
33. Coulter I, Frewin S, Krassas GE, et al. Psychological implications of Graves' orbitopathy. *Eur J Endocrinol*. 2007;157:127–131.
34. In H, Pearce EN, Wong AK, et al. Treatment options for Graves disease: a cost-effectiveness analysis. *J Am Coll Surg*. 2009;209:170–179.e1-2.
35. Wang J, Qin L. Radioiodine therapy versus antithyroid drugs in Graves' disease: a meta-analysis of randomized controlled trials. *Br J Radiol*. 1064;89:20160418.
36. Wilhelm SM, McHenry CR. Total thyroidectomy is superior to subtotal thyroidectomy for management of Graves' disease in the United States. *World J Surg*. 2010;34:1261–1264.
37. Sundaresh V, Brito JP, Wang Z, et al. Comparative effectiveness of therapies for Graves' hyperthyroidism: a systematic review and network meta-analysis. *J Clin Endocrinol Metab*. 2013;98:3671–3677.
38. Mijnhout GS, Franken AA. Antithyroid drug regimens before and after 131I-therapy for hyperthyroidism: evidence-based? *Neth J Med*. 2008;66:238–241.
39. Metso S, Jaatinen P, Huhtala H, et al. Increased cardiovascular and cancer mortality after radioiodine treatment for hyperthyroidism. *J Clin Endocrinol Metab*. 2007;92:2190–2196.
40. Kitahara CM, Berrington de Gonzalez A, Bouville A, et al. Association of radioactive iodine treatment with cancer mortality in patients with hyperthyroidism. *JAMA Intern Med*. 2019;179:1034–1042.
41. Boelaert K, Maisonneuve P, Torlinska B, et al. Comparison of mortality in hyperthyroidism during periods of treatment with thionamides and after radioiodine. *J Clin Endocrinol Metab*. 2013;98:1869–1882.
42. Vestergaard P, Rejnmark L, Weeke J, et al. Smoking as a risk factor for Graves' disease, toxic nodular goiter, and autoimmune hypothyroidism. *Thyroid*. 2002;12:69–75.