Critical Review

The Reality of Randomized Controlled Trials for Assessing the Benefit of Proton Therapy: Critically Examining the Intent-to-Treat Principle in the Presence of Insurance Denial

Mike Hernandez, MS,[a],* J. Jack Lee, PhD,[a] Beow Y. Yeap, ScD,[b] Rong Ye, MS,[a] Robert L. Foote, MD,[c] Paul Busse, MD, PhD,[d] Samir H. Patel, MD,[e] Roi Dagan, MD,[f] James Snider, MD,[g] Nasiruddin Mohammed, MD, MBA,[h] Alexander Lin, MD,[i] Pierre Blanchard, MD, PhD,[j,l] Scott B. Cantor, PhD,[k] Menna Y. Teferra, MS,[l] Kate Hutcheson, PhD,[l,m] Pablo Yepes, PhD,[n,o] Radhe Mohan, PhD,[o] Zhongxing Liao, MD,[l] Thomas F. DeLaney, MD,[d] and Steven J. Frank, MD[l]

[a]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas; [b]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts; [c]Department of Radiation Oncology, Mayo Clinic, Rochester, Minnesota; [d]Department of Radiation Oncology, Massachusetts General Hospital, Boston, Massachusetts; [e]Department of Radiation Oncology, Mayo Clinic, Phoenix, Arizona; [f]Department of Radiation Oncology, University of Florida Health, Gainesville, Florida; [g]Department of Radiation Oncology, University of Maryland Medical System, Baltimore, Maryland; [h]Department of Radiation Oncology, Northwestern Medicine, Warrenville, Illinois; [i]Department of Radiation Oncology, University of Pennsylvania, Philadelphia, Pennsylvania; [j]Department of Radiation Oncology, Gustave Roussy Cancer Campus, Villejuif, France; [k]Department of Health Services Research, The University of Texas MD Anderson Cancer Center, Houston, Texas; [l]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas; [m]Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, Texas; [n]Department of Physics and Astronomy, Rice University, Houston, Texas; and [o]Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas

Abstract

Purpose: This study hypothesized that insurance denial would lead to bias and loss of statistical power when evaluating the results from an intent-to-treat (ITT), per-protocol, and as-treated analyses using a simulated randomized clinical trial comparing proton therapy to intensity modulated radiation therapy where patients incurred increasing rates of insurance denial.

Methods and Materials: Simulations used a binary endpoint to assess differences between treatment arms after applying ITT, per-protocol, and as-treated analyses. Two scenarios were developed: 1 with clinical success independent of age and another assuming dependence on age. Insurance denial was assumed possible for patients <65 years. All scenarios considered an age distribution with mean $\pm$ standard deviation: 55 ± 15 years, rates of insurance denial ranging from 0%-40%, and a sample of N = 300 patients (150 per arm). Clinical success rates were defined as 70% for proton therapy and 50% for intensity modulated radiation therapy. The average treatment effect, bias, and power were compared after applying 5000 simulations.

Results: Increasing rates of insurance denial demonstrated inherent weaknesses among all 3 analytical approaches. With clinical success independent of age, a per-protocol analysis demonstrated the least bias and loss of power. When clinical success was dependent on age, the per-protocol and ITT analyses resulted in a similar trend with respect to bias and loss of power, with both outperforming the as-treated analysis.

Conclusions: Insurance denial leads to misclassification bias in the ITT analysis, a missing data problem in the per-protocol analysis, and covariate imbalance between treatment arms in the as-treated analysis. Moreover, insurance denial forces the critical appraisal of patient features (eg, age) affected by the denial and potentially influencing clinical success. In the presence of insurance denial, our study suggests cautious reporting of ITT and as-treated analyses, and placing primary emphasis on the results of the per-protocol analysis.

Introduction

Randomized controlled trials (RCTs) provide level 1 evidence for establishing standards of care within the National Comprehensive Cancer Network guidelines.[1] Proton therapy is considered an advanced form of delivering radiation compared with conventional radiographs, underscoring the need for evidence from RCTs. The past 2 decades have advanced the delivery of both protons (eg, intensity modulated proton therapy) and conventional radiographs (intensity modulated radiation therapy [IMRT]).[2,3] However, providing proton therapy is costly,[4,5] and insurance companies are reluctant to incur higher expenses without proven clinical benefit. This has led to insurance denials for patients randomly assigned to proton therapy.[4-6]

As an example, support for treating oropharyngeal tumors with proton therapy is increasing due to dosimetric advantages that spare normal tissue from radiation exposure.[6-9] Although patients who qualify for Medicare rarely experience insurance denial,[10] younger patient populations, such as those with oropharyngeal tumors, remain vulnerable to the possibility of insurance denial when randomly assigned to proton therapy.[11-13] Although unfortunate, this situation illustrates how insurance denial affects study features, such as age, potentially restricting younger patients from receiving superior treatment. Moreover, this infringes well-established analytical techniques from providing valid estimates when evaluating an RCT. Specifically, the intent-to-treat (ITT) principle, commonly interpreted as "once randomized, always analyzed"[14,15] provides a stringent framework for

evaluating drug trials, where noncompliance and dropouts due to therapeutic administration are expected. However, in radiation therapy, the prescribed dose is normally beyond a patient's control, making it more likely that a patient will complete treatment rather than exiting the study early.

The ITT approach is supported by Food and Drug Administration and National Institutes of Health guidelines for the primary reporting of results from RCTs.[16] The ITT principle is characterized by the following: (1) all randomized subjects must be included in the analysis, (2) participants are assigned to treatment groups regardless of whether they receive their assigned treatment, and (3) all outcomes are observed and included, regardless of their relationship to treatment.[14-18] To visually explain ITT, per-protocol, and as-treated principles, Pearl[19] used a diagram (Fig. 1) to outline the process from random assignment (Z) to receipt of treatment (X), and ultimately to the outcome observed (Y). Diagramming this sequence from randomization to clinical outcome helps to explain the subtleties intrinsic to each analytical approach.[16] Notably, the treatment assignment (Z) through randomization does not directly affect the observed outcome (Y). Instead, treatment assignment (Z) affects the outcome observed (Y) through the treatment received (X). Hence, an ITT analysis can be viewed as discounting X and focusing directly on the assumed relationship Z $\rightarrow$ Y.

Alternatively, a per-protocol analysis is considered a subset analysis consisting of patients who were both randomized and treated according to protocol. This implies that the relationship Z $\rightarrow$ X $\rightarrow$ Y is strictly maintained, discarding patients who fail to receive X, despite

Per-Protocol Assessment

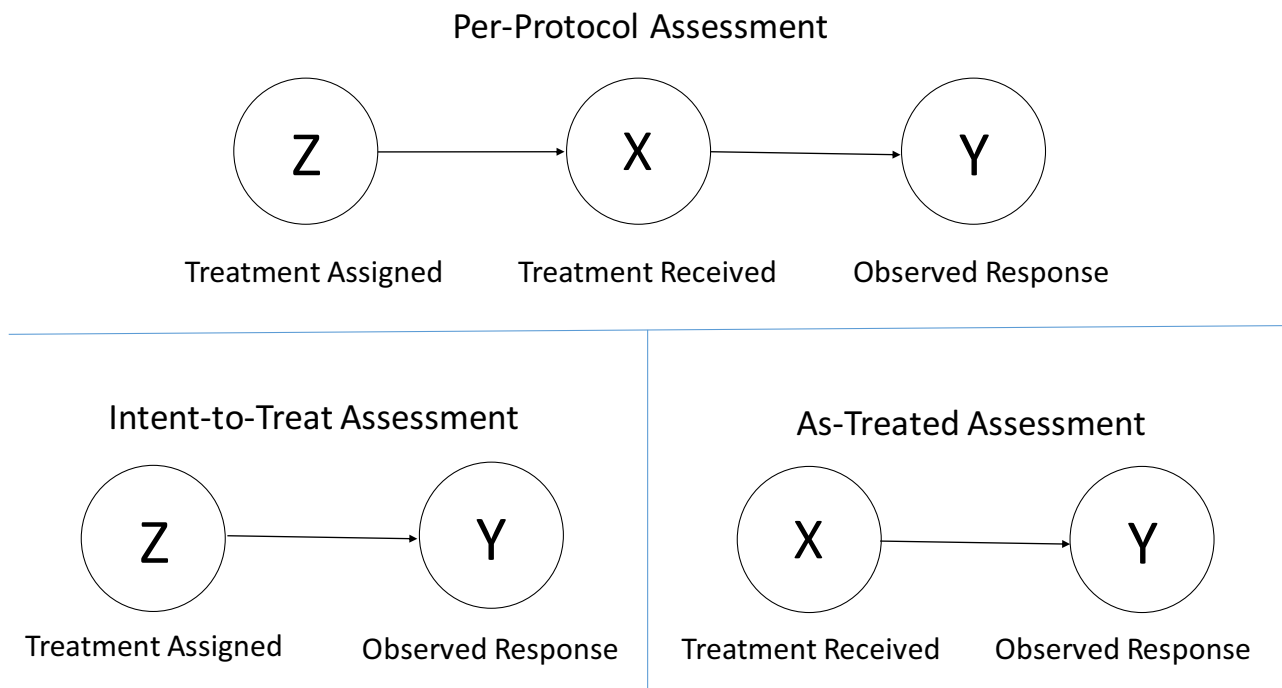## Intent-to-Treat Assessment   As-Treated Assessment

Figure 1   Graphical representation of a randomized trial.

assignment Z. Finally, an as-treated analysis considers patients by their received treatment, rather than their randomized assignment. Applying an as-treated analysis to data from a clinical trial focuses solely on the X → Y relationship, discounting Z.[19-21] Therefore, this study hypothesized that insurance denial would lead to bias and loss of power when evaluating the results of ITT, per-protocol, and as-treated analyses.

Methods and Materials

Our simulation study established fixed parameters considering a binary outcome to describe clinical success. The following assumptions were stipulated: (1) a fixed difference exists in the rate of clinical success between proton therapy and IMRT and (2) proton therapy is superior in outcome by 20 percentage points. Therefore, we chose a 70% clinical success rate for proton therapy and 50% for IMRT and a total sample size of N = 300 (150 receiving proton therapy and 150 IMRT). This study parameterization was used intentionally to confer 94% power with the use of a 2-sided χ^2 test of equal proportions given an alpha-level of 5%. Overpowering was required to establish 80% statistical power as an acceptable threshold.

To align our simulations with reality, younger patients (those <65 years of age) were considered prone to insurance denial, given that patients ≥65 years normally qualify for Medicare coverage. Two scenarios were

developed: one assumed that clinical success was independent of age and another where clinical success was dependent on age. The same age distribution (mean ± standard deviation, 55 ± 15 years) and rates of insurance denial (range, 0%-40%) were used across all scenarios.

Clinical success rates were estimated for proton therapy (Pe) and IMRT (Ps), whereas π_e and π_s were used to denote population parameters. The estimated treatment effect of 20% was computed by taking the difference between clinical success rates of proton therapy and IMRT ($Pe - Ps$). Bias was computed by taking the difference between ($Pe - Ps$) and ($\pi_e - \pi_s$). The test statistic

$$zTS = \frac{(Pe - Ps)}{SE}, \text{ where}$$

$$SE = \sqrt{\frac{Pe(1 - Pe)}{N Protons} + \frac{Ps(1 - Ps)}{N IMRT}}$$

was used to derive $P(Z > zTS \mid H_0$ is true), the P value indicating the probability of observing a test statistic as extreme as zTS under the null hypothesis of ($\pi_e - \pi_s = 0$).[22] Study sample sizes for proton therapy (N$Protons$) and IMRT (N$IMRT$) were scenario dependent, given the analytical method used and the rate of insurance denial incurred.

In applying each principle (ITT, per-protocol, and as-treated) to the scenario of clinical success and age dependence, the overall success rates remained consistent

Table 1   Success rates of proton therapy as a function of insurance denial and age distribution analyzed under the ITT principle

Rate of insurance denial	Average treatment effect (2.5th and 97.5th percentiles)	Average bias (2.5th and 97.5th percentiles)	Average *P* values (2.5th and 97.5th percenttiles)	Statistical power (% *P* < .05)
Clinical success independent of age (ITT scenario 1)				
0%	0.200 (0.087, 0.307)	0.000 (−0.113, 0.107)	.014 (.000, .150)	0.932
5%	0.193 (0.080, 0.300)	0.007 (−0.120, 0.100)	.019 (.000, .188)	0.913
10%	0.185 (0.073, 0.293)	0.015 (−0.127, 0.093)	.027 (.000, .240)	0.890
15%	0.179 (0.073, 0.287)	0.021 (−0.127, 0.087)	.030 (.000, .248)	0.867
20%	0.171 (0.060, 0.280)	0.029 (−0.140, 0.080)	.039 (.000, .344)	0.835
25%	0.163 (0.053, 0.273)	0.037 (−0.147, 0.073)	.050 (.000, .412)	0.788
30%	0.155 (0.047, 0.260)	0.045 (−0.153, 0.060)	.060 (.000, .478)	0.757
35%	0.148 (0.033, 0.260)	0.052 (−0.167, 0.060)	.081 (.000, .643)	0.710
40%	0.141 (0.027, 0.247)	0.059 (−0.173, 0.047)	.091 (.000, .725)	0.668
Clinical success dependent on age (ITT scenario 2)				
0%	0.200 (0.093, 0.307)	0.000 (−0.107, 0.107)	.015 (.000, .127)	0.934
5%	0.193 (0.087, 0.300)	0.007 (−0.113, 0.100)	.018 (.000, .162)	0.921
10%	0.184 (0.073, 0.287)	0.016 (−0.127, 0.087)	.026 (.000, .239)	0.887
15%	0.178 (0.073, 0.287)	0.022 (−0.127, 0.087)	.030 (.000, .242)	0.865
20%	0.171 (0.060, 0.280)	0.029 (−0.140, 0.080)	.038 (.000, .345)	0.834
25%	0.163 (0.053, 0.267)	0.037 (−0.147, 0.067)	.051 (.000, .416)	0.795
30%	0.156 (0.040, 0.267)	0.044 (−0.160, 0.067)	.063 (.000, .544)	0.758
35%	0.147 (0.040, 0.260)	0.053 (−0.160, 0.060)	.077 (.000, .561)	0.704
40%	0.139 (0.033, 0.253)	0.061 (−0.167, 0.053)	.094 (.000, .641)	0.662

Abbreviation: IMRT = intensity modulated radiation therapy; ITT = intent-to-treat.
Each treatment arm consisted of 150 patients (NProtons = 150 and NIMRT = 150). Statistical power was defined as % *P* value < .05. "Age distribution" refers to a mean age of 55 ± 15 years, and the 2.5th and 97.5th percentiles were derived from empirical distributions.

with our previous definitions (70% for proton therapy vs 50% for IMRT); however, for proton therapy, the conditional success rate characterized by the probability of clinical success in patients <65 years, $Prob(ClinicalSuccess|Age <65)$, was inflated to 80%. These simulations were structured to provide an advantage to younger participants treated with proton therapy while maintaining the marginal clinical success rate of 70% for proton therapy. Finally, the empirical power for each study scenario was derived using the proportion of *P* values emanating from 5000 simulated RCTs, providing a statistically significant result (ie, *P* < .05). The average treatment effect, bias, and study power were compared across different analytical approaches and between study scenarios.

Results

The results of the simulation study for each analytical approach (ITT, per-protocol, and as-treated) are described in the following sections.

Intent-to-treat analysis: Misclassification bias

Applying the ITT principle to a simulated RCT involving insurance denial revealed how misclassification biased the treatment effect. First, the ITT principle was applied to the simulations in which clinical success and age were independent, so clinical success remained the same regardless of patient age. ITT focuses on the Z → Y relationship, leading to misclassification in the Z → X (ie, the randomization and treatment assignment) relationship with the occurrence of insurance denial. Because ITT ignores treatment assignment, insurance denial of proton therapy had no effect, post-randomization, on the sample size per treatment arm (NProtons = 150 and NIMRT = 150). After 5000 RCT simulations with insurance denial increasing from 0% to 40%, it became clear that allowing patients randomized to proton therapy to receive IMRT diluted the success rate of proton therapy (Table 1). Results further demonstrated that once insurance denial reached 25% in the proton therapy arm, the misclassification bias introduced with an ITT analysis reduced the treatment effect from 20% to 16.3%, and empirical power fell from 93% to 79%.

The next simulation study explored the effect of using an ITT analysis to evaluate the treatment effect of proton therapy, assuming that clinical success depended on age. Again, the sample size per treatment arm was not affected; however, as the rate of insurance denial increased, the clinical success rate of proton therapy decreased due to misclassification bias (Table 1). After 5000 simulations, the average bias incurred as a function of insurance denial was similar to the bias incurred if clinical success and age were independent; therefore,
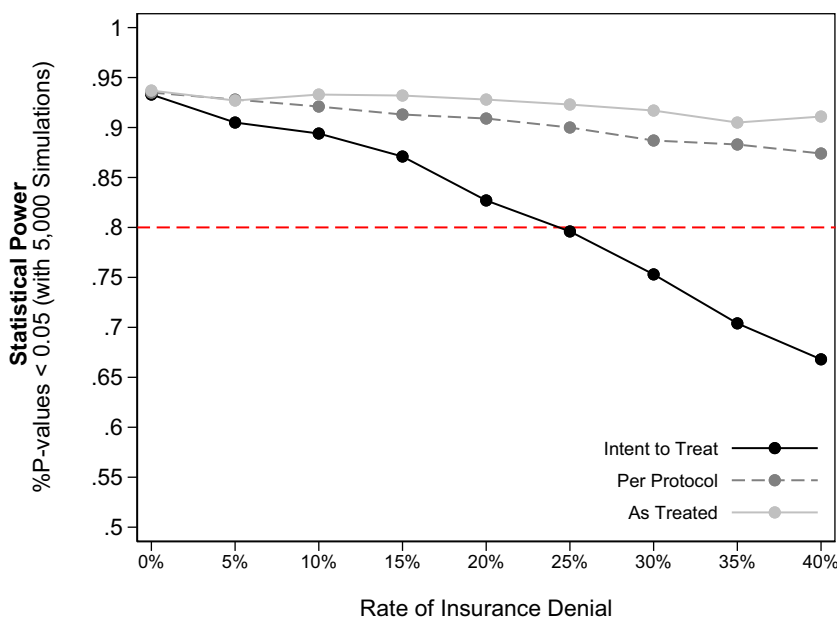
Figure 2   Trends in statistical power assuming independence between age and clinical success.

regardless of the dependency of clinical success on age, the bias from an ITT analysis remained consistent. When insurance denial exceeded 25%, empirical power deteriorated below 80% (Table 1).

Per-protocol analysis: The missing data problem

When the per-protocol principle was applied, insurance denial caused a missing data problem. Adherence to the $Z \rightarrow X \rightarrow Y$ relationship required eliminating the subset of patients who were randomized to receive proton therapy but who were not treated accordingly due to insurance denial. In this case, applying a per-protocol analysis across 5000 simulations in which clinical success and age were assumed independent resulted in a diminishing sample size occurring in parallel with the rate of insurance denial within the proton therapy arm. When the rate of insurance denial reached 40%, the average sample size was reduced to N$Protons$ = 90. Notably, statistical power never fell below the 80% threshold with the highest explored insurance denial rate (40%; see Fig. 2). This was expected for this scenario, in which the missing data were considered "ignorable" and a random subset of the original study arm sample. However, when clinical success was assumed to depend on age, the missing data in the per-protocol analysis became "nonignorable" because patients <65 (with a higher propensity for clinical success) were being discarded from the analysis set. As the insurance denial rate reached 25% and 40%, the empirical power fell from 93.4% to 81.3% and 93.4% to 67%, respectively (Table 2).

As-treated analysis: Covariate imbalance

The as-treated analysis focused strictly on the relationship $X \rightarrow Y$, ignoring Z random assignment. In the first scenario, age and clinical success were assumed independent. With an as-treated analysis, the total sample size (eg, N$Total$ = 300) can be maintained as long as patients are followed to endpoint Y after breaching $Z \rightarrow X$. Under this assumption, the total sample size is maintained (N = 300), but insurance denial imbalanced the sample sizes and age distribution of the treatment arms, as only proton therapy patients were denied insurance and were then treated with IMRT. As the rate of insurance denial increased, treatment effect estimates became less precise but remained relatively unbiased (Table 3). However, assuming dependence between age and clinical success, the imbalance in the sample sizes and the age distribution of the treatment arms affected participants <65 years with a higher propensity for clinical success from proton therapy, but they were ultimately treated with IMRT. In this scenario, as the rate of insurance denial reached 25%, the treatment effect experienced severe bias and empirical power fell from 93.7% to 75.4% (Table 3).

Misclassification bias, missing data, and covariate imbalance

Relative comparisons were made across analytical methods using empirical power after generating 5000 simulated RCTs. In the scenario assuming independence between age and clinical success, the change in empirical power as a function of insurance denial is illustrated in Figure 2. Notably, when the insurance denial rate was 0%,

Table 2   Success rates of proton therapy as a function of insurance denial and age distribution as analyzed under the per-protocol principle

Rate of insurance denial	Average treatment effect (2.5th and 97.5th percentiles)	Average bias (2.5th and 97.5th percentiles)	Average P values (2.5th and 97.5th percentiles)	Statistical power (% $P < .05$)
Clinical success independent of age (per-protocol scenario 1)				
0%	0.199 (0.087, 0.307)	0.001 (−0.113, 0.107)	.015 (.000, .153)	0.930
5%	0.200 (0.089, 0.311)	0.000 (−0.111, 0.111)	.016 (.000, .150)	0.931
10%	0.199 (0.086, 0.309)	0.001 (−0.114, 0.109)	.017 (.000, .165)	0.919
15%	0.199 (0.090, 0.307)	0.001 (−0.110, 0.107)	.017 (.000, .152)	0.915
20%	0.199 (0.088, 0.309)	0.001 (−0.112, 0.109)	.019 (.000, .172)	0.911
25%	0.200 (0.085, 0.313)	0.000 (−0.115, 0.113)	.021 (.000, .193)	0.904
30%	0.201 (0.086, 0.316)	0.001 (−0.114, 0.116)	.023 (.000, .200)	0.885
40%	0.202 (0.081, 0.319)	0.002 (−0.119, 0.119)	.028 (.000, .246)	0.877
Clinical success dependent on age (per-protocol scenario 2)				
0%	0.201 (0.093, 0.307)	0.001 (−0.107, 0.107)	.014 (.000, .124)	0.934
5%	0.196 (0.087, 0.307)	0.004 (−0.113, 0.107)	.019 (.000, .160)	0.916
10%	0.192 (0.079, 0.305)	0.008 (−0.121, 0.105)	.023 (.000, .210)	0.895
15%	0.188 (0.079, 0.295)	0.012 (−0.121, 0.095)	.025 (.000, .219)	0.879
20%	0.182 (0.070, 0.291)	0.018 (−0.130, 0.091)	.034 (.000, .286)	0.840
25%	0.177 (0.063, 0.292)	0.023 (−0.137, 0.092)	.042 (.000, .355)	0.813
30%	0.171 (0.054, 0.286)	0.029 (−0.146, 0.086)	.056 (.000, .448)	0.766
35%	0.165 (0.043, 0.282)	0.035 (−0.157, 0.082)	.069 (.000, .569)	0.728
40%	0.158 (0.036, 0.274)	0.042 (−0.164, 0.074)	.087 (.000, .648)	0.670

Abbreviation: IMRT = intensity modulated radiation therapy.

Each treatment arm initially consisted of 150 patients (NProtons = 150 and NIMRT = 150); however, when the rate of insurance denial reached 40%, the mean number in the proton therapy arm was reduced to 90 patients.

Statistical power was defined as % P value < .05.

"Age distribution" refers to a mean age of 55 ± 15 years, and the 2.5th and 97.5th percentiles were derived from empirical distributions.

all of the analytical methods remained powered at approximately 94%, as confirmed by a power analysis with standard statistical software (nQuery Advisor v7), which was used to provide sample size justification. Once insurance denial reached 25%, the separation between trends illustrating loss of power was evident. However, because the study had been overpowered to detect a 20 percentage point increase in clinical success for proton therapy, an insurance denial rate of 25% maintained a viable threshold with respect to empirical power (>80%).

In another scenario, relative comparisons assuming dependence of age and clinical success revealed that all of the analytical approaches were deficient in providing unbiased estimates of the treatment effect as the rate of insurance denial increased (Fig. 3). When the rate reached 25%, only the per-protocol analysis maintained adequate empirical power. The as-treated analysis had, as expected, the most severe deterioration in power, reinforcing the need to use appropriate methods to adjust for the treatment effect.

Discussion

Our simulation study provided insight into the consequences of using ITT, per-protocol, and as-treated

analyses when evaluating an RCT to investigate the treatment effect of proton therapy versus IMRT in the presence of increasing rates of insurance denial. Specifically, insurance denial led to misclassification bias in the ITT analysis, a missing data problem in the per-protocol analysis, and covariate imbalance between treatment arms in the as-treated analysis. Scenarios introducing independence of clinical success from age and dependence of clinical success on age further exposed the need to explore baseline patient characteristics for their association with both insurance denial and clinical success.

In the context of drug trials, the ITT principle forces patients dropping out from negative treatment effects to remain analyzable. Retaining only patients who tolerate, comply, and complete their prescribed study regimen while eliminating patients experiencing poor outcomes increases the likelihood that a treatment arm would appear more beneficial. The ITT principle protects against this type of bias by assigning all randomized patients an outcome. The advantage of treating study dropouts as failures, for example, is that it penalizes a treatment arm for losing patient information while maintaining the study's original sample size.[10,18,21]

Although the ITT analysis seems appealing, we demonstrated how it can also be problematic. The increasing rates of insurance denial proportionately

Table 3   Success rates of proton therapy as a function of insurance denial and age distribution as analyzed under the as-treated principle

Rate of insurance denial	Average treatment effect (2.5th and 97.5th percentiles)	Average bias (2.5th and 97.5th percentiles)	Average P values (2.5th and 97.5th percentiles)	Statistical power (% $P < .05$)
Clinical success independent of age (as-treated scenario 1)				
0%	0.200 (0.093, 0.307)	0.000 (−0.107, 0.107)	.014 (.000, .133)	0.935
5%	0.199 (0.091, 0.308)	0.001 (−0.109, 0.108)	.015 (.000, .133)	0.932
10%	0.201 (0.094, 0.308)	0.001 (−0.106, 0.108)	.014 (.000, .122)	0.935
15%	0.199 (0.092, 0.303)	0.001 (−0.108, 0.103)	.016 (.000, .138)	0.927
20%	0.201 (0.092, 0.308)	0.001 (−0.108, 0.108)	.016 (.000, .139)	0.927
25%	0.200 (0.090, 0.305)	0.000 (−0.110, 0.105)	.016 (.000, .155)	0.927
30%	0.200 (0.088, 0.312)	0.000 (−0.112, 0.112)	.019 (.000, .169)	0.92
35%	0.198 (0.086, 0.309)	0.002 (−0.114, 0.109)	.020 (.000, .181)	0.906
40%	0.199 (0.086, 0.309)	0.001 (−0.114, 0.109)	.021 (.000, .193)	0.912
Clinical success dependent on age (as-treated scenario 2)				
0%	0.200 (0.093, 0.307)	0.000 (−0.107, 0.107)	.014 (.000, .126)	0.937
5%	0.193 (0.080, 0.303)	0.007 (−0.120, 0.103)	.021 (.000, .197)	0.913
10%	0.185 (0.077, 0.295)	0.015 (−0.123, 0.095)	.024 (.000, .212)	0.89
15%	0.177 (0.064, 0.281)	0.023 (−0.136, 0.081)	.033 (.000, .312)	0.856
20%	0.169 (0.056, 0.279)	0.031 (−0.144, 0.079)	.044 (.000, .396)	0.812
25%	0.160 (0.044, 0.272)	0.040 (−0.156, 0.072)	.060 (.000, .517)	0.754
30%	0.153 (0.042, 0.264)	0.047 (−0.158, 0.064)	.072 (.000, .548)	0.711
35%	0.144 (0.033, 0.258)	0.056 (−0.167, 0.058)	.093 (.000, .640)	0.646
40%	0.135 (0.017, 0.248)	0.065 (−0.183, 0.048)	.125 (.000, .810)	0.568

Abbreviation: IMRT = intensity modulated radiation therapy.

Each treatment arm initially consisted of 150 patients (N*Protons* = 150 and N*IMRT* = 150); however, when clinical success was assumed independent of age, the treatment effect became less precise while remaining relatively unbiased as the rate of insurance denial increased. When clinical success was assumed to be dependent on age and once the insurance denial rate reached 25%, the treatment effect became severely biased and statistical power fell to below 80%. Statistical power was defined as % P value < .05.

"Age distribution" refers to a mean age of 55 ± 15 years, and the 2.5th and 97.5th percentiles were derived from empirical distributions.
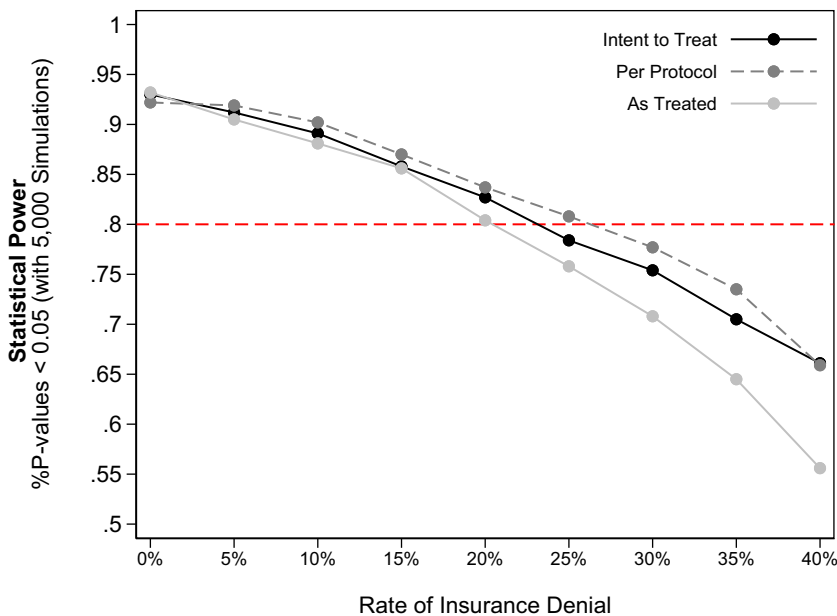

Figure 3   Trends in statistical power assuming dependence between age and clinical success.

increased the number of patients who were assigned to protons but actually received IMRT (ie, misclassified patients). Deliberately including misclassified patients in the analysis hampered the treatment benefit of protons while it augmented the treatment benefit of IMRT. With increasing rates of insurance denial, the minimum detectable difference between treatment arms was lessened. This consequently diminished the study's power to declare a result statistically significant. The extent of misclassification bias introduced by the ITT analysis relative to per-protocol and as-treated analyses was most apparent in the scenario where clinical success was independent of age.

With insurance denial and age inextricably linked, examining baseline covariates of an RCT is required to determine whether imbalances in patient characteristics exist due to insurance denial. Analytical remedies for addressing covariate imbalance in RCTs have been proposed by Lee et al.[23], who discussed the shortcomings arising from the disruption of the randomization process that led to reduced sample sizes, biased estimates, and invalid statistical tests. They cautioned against ignoring the randomization process and the use of an as-treated analysis. Ultimately, the authors considered the ITT analysis as the primary analysis to be reported.

The idea that an ITT analysis has the potential to yield biased estimates is not new. Sheiner and Rubin[17] drew this conclusion in their critique of the ITT analysis. After making a theoretical comparison with per-protocol and as-treated analyses, they suggested the use of model-based methods to acquire statistically valid estimates of the treatment effect. Hernan and Hernandez-Diaz[20] compared ITT, per-protocol, and as-treated analyses, and ultimately advocated using all 3. The ITT analysis was considered the primary analysis to be reported, followed by the per-protocol analysis, and finally the as-treated analysis with appropriate adjustment for confounding.

An as-treated analysis considers patients by the actual treatment received rather than assigned via randomization. In the scenario where clinical success was dependent on age, the ITT and per-protocol analyses performed similarly, and both slightly outperformed the as-treated analysis. An as-treated analysis disrupts the randomization process, requiring investigators to monitor for imbalances in baseline patient characteristics. Balance in baseline covariates between treatment arms can be attained using post-randomization techniques such as matching, weighting, or implementing regression modeling strategies, all well-described in the literature.[24-29] Although these methods enhance the reliability of study results, Sheiner and Rubin[17] warned against using regression models without appropriately diagnosing joint distributions where substantial confounding may be present. As an example, consider a situation where all patients younger than 65 years of age are denied insurance. This

would be detrimental for any analysis due to insufficient overlap between treatment arms on a key patient characteristic (eg, age). Moreover, this situation would impede regression models from using age as a covariate to adjust the treatment effect. Applying an as-treated analysis to data from a clinical trial has pitfalls.[23-26] Sample sizes between treatment arms become imbalanced and the potential for confounding increases.[27-30] Thus, results from as-treated analyses should only be reported after using covariate adjustment, as described by Lee et al.[23,28-30]

The ITT, per-protocol, and as-treated approaches all have inherent weaknesses in the presence of insurance denial. However, a per-protocol analysis may be the only reliable alternative to an ITT or as-treated analysis, with the caveat that a missing data problem arises. The per-protocol analysis includes only patients randomized and treated as described by a clinical protocol. In our simulation study, this analytical approach excluded patients randomized to proton therapy who received IMRT due to insurance denial. Excluding these patients not only reduced the sample size in the proton therapy arm, it also eliminated more patients >65 years. If the missing data arising from a per-protocol analysis is considered random, then the remaining patient information is simply a random sample of the original study's sample size. Although the reduction in sample size diminishes statistical power and estimation precision, the treatment effect remains unbiased. With increasing rates of insurance denial, the per-protocol analysis outperformed the ITT analysis in the scenario where clinical success was independent of age, and it outperformed the as-treated analysis in the scenario where clinical success was dependent on age. Therefore, our study suggests cautious reporting of ITT and as-treated analyses in the presence of insurance denial and placing primary emphasis on the results of the per-protocol analysis.

Acknowledgments

References

1. National Comprehensive Cancer Network (NCCN) Guidelines. Available at: https://www.nccn.org/professionals/physician_gls/default.aspx. Accessed April 4, 2020.

2. Kandula S, Zhu X, Garden S, et al. Spot-scanning proton beam therapy vs intensity modulated radiation therapy for ipsilateral head and neck malignancies: A treatment planning comparison. *Med Dosim*. 2013;38:390-394.

3. Frank SJ, Cox JD, Gillin M, et al. Multifield optimization intensity modulated proton therapy for head and neck tumors: A translation to practice. *Int J Rad Onc Bio Phys*. 2014;89:846-853.

4. Bekelman JE, Denicoff A, Buchsbaum J. Randomized trials of proton therapy: Why they are at risk, proposed solutions, and implications for evaluating advanced technologies to diagnose and treat cancer. *J Clin Oncol*. 2018;36:2461-2464.

5. Mailhot Vega RB, Ishaq O, Raldow A, et al. establishing cost-effective allocation of proton therapy for breast irradiation. *Int J Radiat Oncol Biol Phys*. 2016;95:11-18.

6. Frank SJ, Blanchard P, Lee JJ, et al. Comparing intensity-modulated proton therapy with intensity- modulated photon therapy for oropharyngeal cancer: The journey from clinical trial concept to activation. *Semin Radiat Oncol*. 2018;28:108-113.

7. Holliday EB, Kocak-Uzel E, Feng L, et al. Dosimetric advantages of intensity-modulated proton therapy for oropharyngeal cancer compared with intensity-modulated radiation: A case-matched control analysis. *Med Dosim*. 2016;41:189-194.

8. Blanchard P, Garden S, Gunn GB, et al. Intensity-modulated proton beam therapy (IMPT) versus intensity-modulated photon therapy (IMRT) for patients with oropharynx cancer − A case matched analysis. *Radiother Oncol*. 2016;120:48-55.

9. Sio TT, Lin H, Shi Q, et al. Intensity modulated proton therapy versus intensity modulated photon radiation therapy for oropharyngeal cancer: First comparative results of patient-reported outcomes. *Int J Rad Oncol Biol Phys*. 2016;95:1107-1114.

10. Does Medicare Cover Proton Therapy? Available at: www.medicare.org/articles/does-medicare-cover-proton-therapy. Accessed October 2, 2020.

11. Goitein M, Cox JD. Should randomized clinical trials be required for proton radiotherapy? *J Clin Oncol*. 2008;26:175-176.

12. Bentzen SM. Randomized controlled trials in health technology assessment: Overkill or overdue? *Radiother Oncol*. 2008;86:142-147.

13. Kong F-M. What happens when proton meets randomization: Is there a future for proton therapy? *J Clin Oncol*. 2018;36:1777-1779.

14. Newell DJ. Intention-to-treat analysis: Implications for qualitative and qualitative research. *Int J Epidemiol*. 1992;21:837-841.

15. Wertz RT. Intent to treat: Once randomized, always analyzed. *Clin Aphasiol*. 1995;23:57-64.

16. Cook TD, DeMets DL. *Introduction to Statistical Methods for Clinical Trials*. Boca Raton, FL: Chapman & Hall-CRC; 2008.

17. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goal of clinical trials. *Clin Pharmacol Ther*. 1995;57:6-15.

18. Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trial*. 2000;21:167-189.

19. Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. New York, NY: Cambridge University Press; 2009.

20. Hernan M, Hernandez-Diaz S. Beyond the intention to treat in comparative effectiveness research. *Clin Trial*. 2012;9:48-55.

21. Ang KK, Zhang Q, Rosenthal D, et al. Randomized phase III trial of concurrent accelerated radiation plus cisplatin with or without cetuximab for stage III to IV head and neck carcinoma: RTOG 0522. *J Clin Oncol*. 2014;32:2940-2950.

22. Forthofer RN, Lee ES, Hernandez M. *Biostatistics: A Guide to Design, Analysis, and Discovery*. 2nd ed. Burlington: Academic Press; 2006.

23. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: Is it really an option? *Stat Med*. 1991;10:1595-1605.

24. Goetghebeur E, Molenberghs G, Katz J. Estimating the causal effect of compliance on binary outcome in randomized control trials. *Stat Med*. 1998;17:341-355.

25. Xie H, Heitjan D. Sensitivity analysis of casual inference in a clinical trial subject to crossover. *Clin Trial*. 2004;1:21-30.

26. Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med*. 2000;19:1849-1864.

27. Senn S. Testing for baseline balance in clinical trials. *Stat Med*. 1994;13:1715-1726.

28. Moher D, Schulz KF, Altman D. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001;285:1987-1991.

29. Austin PC, Manca A, Zwarenstein M, et al. A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: A review of trials published in leading medical journals. *J Clin Epidemiol*. 2010;63:142-153.

30. Roberts C, Torgerson DJ. Understanding controlled trials: Baseline imbalance in randomised controlled trials. *BMJ*. 1999;319:185.