Research Article

Cancer Immunology Research

# Treatment-Related Adverse Events Predict Improved Clinical Outcome in NSCLC Patients on KEYNOTE-001 at a Single Center

Aaron Lisberg, D. Andrew Tucker, Jonathan W. Goldman, Brian Wolf, James Carroll, Ariana Hardy, Karolyn Morris, Paulina Linares, Carlos Adame, Marshall L. Spiegel, Courtney Wells, Jordan McKenzie, Blanca Ledezma, Melody Mendenhall, Phillip Abarca, Krikor Bornazyan, Jaime Hunt, Nima Moghadam, Natalie Chong, Danielle Nameth, Caitlin Marx, John Madrigal, Sitaram Vangala, Narek Shaverdian, David Elashoff, and Edward B. Garon



## Abstract

We retrospectively analyzed non–small cell lung cancer (NSCLC) patients from a single center treated with pembrolizumab on the KEYNOTE-001 trial and evaluated the association between treatment-related adverse events (trAEs) and clinical outcomes. Investigators reported AEs on trial and graded them according to Common Terminology Criteria for Adverse Events v4.0, labeling them as unlikely, possibly, or probably treatment-related. AEs labeled as possibly/probably related were considered trAEs for this analysis. The relationship between the incidence of a trAE and clinical outcomes was evaluated. Ninety-seven NSCLC patients treated on KEYNOTE-001 at the University of California, Los Angeles were evaluated. Ten percent (85/826) of AEs were trAEs, occurring in 40% (39/97) of patients. The most frequent trAEs were rash (21% patients), fatigue (6% patients), and hypothyroidism (6% patients). The 39 patients that experienced a trAE had increased objective response rate (ORR, 38.5%), progression-free survival (PFS: median, 248 days), and overall survival (OS: median, 493 days), compared with the 58 patients that did not (ORR: 8.9%, PFS: median 60 days, OS: median 144.5 days). The observed association between trAEs and improved clinical outcome persisted when using Cox proportional hazards regression models to assess the confounding effect of covariates and mitigate guarantee-time bias. The association also remained when data were substratified by grade, degree of association, and treatment-related select AE designation. This single-center analysis revealed that trAEs predicted for improved clinical outcome with pembrolizumab, and when controlling for guarantee-time bias and plausible confounders, this association remained. This observed relationship adds to our understanding of anti–PD-1 therapy and could aid clinicians in identifying patients most likely to benefit from therapy. *Cancer Immunol Res; 6(3); 288–94. ©2018 AACR.*



## Introduction

Immune-based cancer therapy has been in development for almost four decades (1–3), but the first immunotherapeutics approved for non–small cell lung cancer (NSCLC) were not until 2015, when the programmed cell death-1 (PD-1) inhibitors pembrolizumab and nivolumab were granted FDA approval (4–7). Because immunotherapy can have mechanisms of action different from conventional cytotoxic chemotherapy and targeted therapies, its side effects can vary as well (8–10). Treatment-related adverse events (trAEs) associated with anti–PD-1 therapy are thought to arise as a result of a robust, uninhibited immune response, resulting in primarily T cell–mediated autoimmunity and/or inflammation that is most commonly transient in nature but can be severe in some instances (8–11). Non–organ-specific trAEs, such as fatigue, fevers, and chills, are commonly seen as a result of immune checkpoint inhibition (8–11). Organ-specific side effects can also occur in nearly any location throughout the body, but are commonly observed in the integumentary, gastrointestinal, endocrine, and hepatic systems (8–11). Of particular concern in patients with NSCLC is the occurrence of trAEs that affect the respiratory system, namely pneumonitis, and although these cases are rare, they can be fatal (6, 9).

It has been hypothesized that certain patients can be primed to respond to immunotherapy due to heightened baseline "immunocompetence," and evidence exists to suggest that patients who experience AEs related to immunotherapy are more likely to benefit from therapy (10, 12–14). Importantly, the largest evaluation to date (a pooled analysis of four clinical trials evaluating nivolumab in melanoma) found that treatment-related select AEs of any grade were associated with improved objective response rate (ORR; ref. 15). However, others have not found a relationship between AEs associated with immunotherapy and clinical outcome (16). Because the association between AEs experienced by NSCLC patients on anti–PD-1 clinical trials and

David Geffen School of Medicine at UCLA, Los Angeles, California.

**Note:** Supplementary data for this article are available at Cancer Immunology Research Online (http://cancerimmunolres.aacrjournals.org/).

**Clinical Trials Identifier:** NCT01295827

**Corresponding Author:** Edward B. Garon, David Geffen School of Medicine at UCLA, 2825 Santa Monica Blvd., Suite 200, Santa Monica, CA 90404. Phone: 310-586-2098; Fax: 310-586-0841; E-mail: egaron@mednet.ucla.edu

**doi:** 10.1158/2326-6066.CIR-17-0063

clinical outcome has not formally been evaluated, we pursued a retrospective analysis of clinical data obtained from the 97 NSCLC patients treated at our single center [The University of California, Los Angeles (UCLA)] as part of the KEYNOTE-001 trial and evaluated the association between trAEs and clinical outcomes.

## Materials and Methods

### Study design

We performed a retrospective analysis of 97 NSCLC patients treated at UCLA with follow-up data available between August 2012 and December 2016 for KEYNOTE-001 (clinicaltrials.gov NCT01295827). A 98th patient was enrolled at UCLA but transferred care to another institution shortly after enrollment, so this patient was not included in the analysis because no follow-up data were available. Patients were treated with pembrolizumab intravenously at a dose of either 2 mg/kg every 3 weeks or 10 mg/kg every 2 or 3 weeks. Investigators reported AEs and graded them according to Common Terminology Criteria for Adverse Events v4.0 (CTCAE), labeling them as unlikely, possibly, or probably treatment-related. All AEs that were labeled as possibly or probably related to treatment were considered trAEs for our retrospective analysis, regardless of whether one type of trAE was reported multiple times within a particular patient. Investigators also had the ability to independently deem an AE as an immune-related AE (irAE). We also retrospectively identified treatment-related select AEs, as defined by Weber and colleagues, "AEs with a potential immunologic cause that need frequent monitoring and potential intervention with immune suppression and/or endocrine treatment" (15). A history of prior radiotherapy was defined as a patient having received any radiotherapy for the treatment of NSCLC at any time point prior to the first cycle of pembrolizumab, as previously described (17). Finally, per protocol, patients had thyroid-stimulating hormone (TSH), free thyroxine 4 (fT4), and triiodothyronine (T3) assessed at baseline (prior to initiation of therapy), cycle 2, and every other cycle thereafter (i.e., cycle 4, cycle 6, etc.). These parameters were also assessed at unscheduled visits between cycles and at safety follow-up. Tumor responses were evaluated using investigator-assessed, immune-related response criteria (irRC) initially at 9 weeks (+/– 1 week), with subsequent CT or MRI every 9 weeks (6).

### Treatment outcome groups and efficacy analysis

Available efficacy data as of December 31, 2016, was included in all analyses. Three endpoints were included in the efficacy analysis: (i) ORR, the proportion of patients who achieved a complete response or partial response, determined by irRC; (ii) progression-free survival (PFS), the number of days between the first date of treatment with pembrolizumab and date of clinical progression as determined by the treating investigator, radiographic progression by investigator-evaluated irRC, or death from any cause; and (iii) overall survival (OS), the number of days between first date of treatment with pembrolizumab and date of death from any cause. All efficacy analyses were evaluated in the intent-to-treat population that had follow-up data available ($n = 97$).

### Statistical analysis

Clinical and demographic characteristics were compared using the Fisher exact test for categorical variables. PFS and OS curves were created with the Kaplan–Meier estimate and stratified by a history of a trAE. Kaplan–Meier estimates and comparisons between clinical/demographic characteristics were performed using GraphPad Prism v7 (Graphpad Software, Inc.), wherein all tests were two-sided with $P$ values $< 0.05$ considered statistically significant.

Cox proportional hazards regression models were used to assess the effect of trAEs on survival outcomes (PFS and OS), adjusting for potential confounding factors. In the Cox proportional hazards regression models, the number of trAEs was modeled as a time-varying covariate. To assess the effect of trAEs on ORR, adjusting for potential confounders, a logistic regression model was used. In the logistic regression models, only the 65 patients on trial $\geq$9 weeks were included, and only those trAEs that occurred $\leq$9 weeks after initiating therapy were considered. The number of trAEs was modeled as a time-varying covariate in the Cox proportional hazards regression models, and a landmark approach was employed in the logistic regression models because trAE status was unknown at trial initiation and because these approaches mitigate guarantee-time bias. ORRs were compared with ORs. R version 3.3.2 (www.R-project.org) was used for the Cox proportional hazards and logistic regression models with $P$ values <0.05 considered statistically significant.

## Results

### Description of experienced AEs, trAEs, and treatment-related select AEs

A total of 826 AEs were reported on trial occurring in 97% (94/97) of the patients treated at UCLA who had follow-up data available. Of these 826 AEs, 10% (85/826) were deemed to be trAE (53 possibly treatment-related and 32 probably treatment-related), occurring in 40% (39/97) of the UCLA patients. Thirty-three patients experienced a grade 1 trAE, 18 experienced a grade 2 trAE, and 3 experienced a grade $\geq$3 trAE. In contrast, 70.9% (351/495) of the 495 total NSCLC patients on trial experienced a trAE (8). Eleven irAEs were reported in the UCLA cohort, 9 of which were deemed to be treatment-related, and these occurred in 6 patients, 3 with pneumonitis and 3 with rash. Data regarding the frequency of treatment-related irAE occurrence in the 495 total NSCLC patients on trial have not been published to date; 55.3% (47/85) of the prospectively reported trAEs at UCLA were retrospectively identified as treatment-related select AEs, as previously defined (15), and occurred in 28.9% (28/97) of patients. Of the 97 patients treated at UCLA with follow-up data available, 95 were enrolled by one of two investigators. Investigator 1 enrolled 52 patients on trial, of which 39% (20/52) had a trAE, whereas investigator 2 enrolled 43 patients on trial, of which 42% (18/43) had a trAE.

The most frequent AEs experience by UCLA patients on trial were fatigue (50% of patients), pain (36%), and dyspnea (29%; Supplementary Table S1), whereas the most frequent trAEs were rash (21%), fatigue (6%), and hypothyroidism (6%; Table 1). When compared with the total study population, inclusive of the 97 patients treated at UCLA, the incidence of treatment-related hypothyroidism was nearly identical in the UCLA cohort. However, treatment-related rash was much more common in the UCLA cohort, which occurred in 21% of patients compared with only 9.7% in the total study population, whereas treatment-related fatigue was less common in the UCLA cohort, occurring in 6% of patients compared with 19.4% of patients in the total study population. Treatment-related pruritus was more common

**Table 1.** trAEs that occurred in at least 2 of the 97 NSCLC patients treated on KEYNOTE-001 at UCLA

| Treatment-related AEs occurring in ≥2 patients of UCLA cohort (*N* = 97) | | | |
|---|---|---|---|
| **AE term** | **Grade 1** | **Grade 2** | **Grade 3 or 4** |
| Rash | 18 (18.6) | 4 (4.1) | 0 |
| Fatigue | 5 (5.2) | 3 (3.1) | 1 (1.0) |
| Hypothyroidism | 1 (1.0) | 5 (5.2) | 0 |
| Fever/chills | 4 (4.1) | 1 (1.0) | 0 |
| Anorexia | 3 (3.1) | 1 (1.0) | 0 |
| Pneumonitis | 1 (1.0) | 3 (3.1) | 2 (2.1) |
| Pruritus | 2 (2.1) | 1 (1.0) | 0 |
| Abdominal pain | 1 (1.0) | 1 (1.0) | 0 |
| Worsening dyspnea | 0 (0.0) | 2 (2.1) | 0 |
| Joint pain | 1 (1.0) | 1 (1.0) | 0 |
| Edema | 2 (2.1) | 0 | 0 |

NOTE: Of note, no grade 5 trAEs were experienced in this patient cohort.

in the total study population, occurring in 10.7% of patients compared with 3% in the UCLA cohort. No treatment-related grade 5 events were noted in the UCLA cohort (Table 1).

### Baseline clinical characteristics according to history of a trAE

Baseline clinical characteristics of the 97 patients evaluated were well-balanced between those that experienced a trAE and those that did not and were similar to the 495 total patients treated on trial at all sites, inclusive of the 97 UCLA patients (Table 2 and Supplementary Table S2). However, the presence of an EGFR mutation was much more common in the cohort of patients treated at UCLA compared with the total study population (31% and 15%, respectively), which coincided with a higher incidence of "never smokers" in the UCLA cohort (44% compared with 25% in the total study population). The UCLA cohort was less heavily pretreated than the total study population (24% of patients with ≥3 previous lines of therapy in the UCLA cohort compared with 42% in the total study population). Patients that experienced a trAE were numerically more likely to be PD-L1$^+$ (82% of patients) and less likely to be EGFR mutation positive (23% of patients) compared with those that did not experience a trAE (72% and 36%, respectively). The 39 patients that experienced a trAE had a much higher average number of AEs (12.5) compared with the 58 patients that did not experience a trAE (5.8), but the average number of a grade 3 or higher AEs was similar in both populations (1.2 and 1.3, respectively).

**Table 2.** Demographics of 97 NSCLC UCLA patients treated on KEYNOTE-001 stratified by history of trAE at any time point

| Characteristic | trAE$^+$ (*N* = 39) | trAE$^-$ (*N* = 58) | *P* value |
|---|---|---|---|
| Age (y) | | | |
| Median | 67 | 65 | |
| Range | 32–78 | 36–83 | |
| Gender, *n* (%) | | | |
| Male | 22 (56%) | 28 (48%) | 0.4093 |
| Female | 17 (44%) | 30 (52%) | |
| Previous therapy lines, *n* (%) | | | |
| 0 | 7 (18%) | 6 (10%) | 0.5967 |
| 1–3 | 23 (59%) | 38 (66%) | |
| ≥3 | 9 (23%) | 14 (24%) | |
| Smoking status, *n* (%) | | | |
| Ever | 24 (62%) | 30 (52%) | 0.4066 |
| Never | 15 (38%) | 28 (48%) | |
| PD-L1 status, *n* (%) | | | |
| Negative | 5 (13%) | 6 (10%) | 0.2116 |
| Positive | 32 (82%) | 42 (72%) | |
| Unknown | 2 (5%) | 10 (17%) | |
| PD-L1 proportion score, *n* (%) | | | |
| Unknown | 11 (28%) | 10 (17%) | 0.5153 |
| ≤1 | 9 (23%) | 12 (21%) | |
| 1–49 | 14 (36%) | 24 (41%) | |
| ≥50 | 5 (13%) | 12 (21%) | |
| Histology, *n* (%) | | | |
| Nonsquamous | 31 (79%) | 47 (81%) | 1 |
| Squamous cell carcinoma | 8 (21%) | 11 (19%) | |
| Targetable mutations, *n* (%) | | | |
| EGFR mutation | 9 (23%) | 21 (36%) | 0.1873 |
| ALK translocation | 0 (0%) | 2 (3%) | 0.5185 |
| Number of AE occurrences per patient, *n* (%) | | | |
| <3 | 1 (3%) | 12 (21%) | |
| 3–7 | 12 (31%) | 32 (55%) | |
| >7 | 26 (67%) | 14 (24%) | <0.0001 |
| Average number of AEs by grade per patient | | | |
| ≥Grade 3 | 1.2 | 1.3 | 0.8614 |

NOTE: *P* values were computed using the Fisher exact test.

### Effects of prior radiotherapy on the occurrence of a trAE

Because our patient cohort was less heavily pretreated than the total study population and we previously showed that patients in our cohort with a history of prior thoracic radiotherapy (TRT) had more overall treatment-related pulmonary toxicity (12.5% vs. 1.4%, $P = 0{\cdot}046$) compared with those without prior TRT (17), we assessed the influence of prior radiotherapy on the development of a trAE of any type. Prior TRT did not clearly influence the occurrence of a trAE, as 50% (12/24) of patients with a history of prior TRT had a trAE compared with 40% (27/73) of patients without prior history of TRT ($P = 0.338$). We next evaluated the effect of any prior radiotherapy on the occurrence of a trAE, and findings were similar—47.6% (20/42) of patients with a history of prior radiotherapy experienced a trAE compared with 34.6% (19/55) of patients without a history of prior radiotherapy ($P = 0.2154$; Supplementary Table S3).

### Clinical outcomes stratified by history of trAEs

In the UCLA cohort of 97 patients with follow-up data available, an ORR of 21%, median PFS of 64 days, and a median OS of 221 days were observed. The ORR in the total study population, 19.4%, was similar to that observed in the UCLA cohort, although both median PFS and median OS were longer in the total study population (111 and 365 days, respectively; ref. 6). In the UCLA cohort, those that experienced a trAE had a higher ORR (38.5%), a longer median PFS (248 days), and longer median OS (493 days) compared with patients that did not experience a trAE (ORR: 8.9%; PFS: median 60 days; and OS: median 144.5 days; Supplementary Fig. S1A). These relationships persisted when further stratifying patients based on degree of trAE attribution (possible/probable) and when the analysis was limited to only include grade 2 or higher trAEs (Supplementary Table S4). When restricting this analysis to include only the 47 treatment-related select AEs experienced by patients on trial, those that experienced a treatment-related select AE had a higher ORR (39.3%), a longer median PFS (248 days), and longer median OS (493 days) compared with patients who did not (ORR: 13%; PFS: median 61 days; and OS: median 158 days; Supplementary Fig. S1B).

### Further evaluation of the association between a trAE and clinical outcomes

To summarily assess the confounding effect of covariates on the observed association between trAEs and improved clinical

outcomes, as well as mitigate guarantee-time bias, statistical models incorporating clinicopathologic covariates were generated. In the Cox proportional hazards regression models, the number of trAEs was incorporated as a time-varying covariate to control for guarantee-time bias. For PFS, the unadjusted Cox proportional hazards regression model revealed a trend toward improved PFS in patients with increasing numbers of trAEs [HR, 0.81; 95% confidence interval (CI), 0.65–1.01; $P = 0.067$]. This relationship became stronger in the adjusted model that took into account known predictors of response to anti–PD-1 therapy in NSCLC (HR, 0.75; 95% CI, 0.56–0.99; $P = 0.043$). For OS, both the unadjusted and adjusted Cox proportional hazards regression models support a strong association between improved OS in patients with increasing numbers of trAEs (unadjusted model: HR, 0.77; 95% CI, 0.63–0.94; $P = 0.011$ and adjusted model: HR, 0.75; 95% CI, 0.58–0.96; $P = 0.021$). Utilizing logistic regression models to assess the association between ORR and the number of trAEs experienced by patients, the adjusted model incorporating PD-L1 proportion score and EGFR mutational status suggested a trend toward improved ORR in patients that experienced trAEs (unadjusted model: HR, 1.27; 95% CI, 0.81–1.99; $P = 0.306$ and adjusted model: HR, 2.02; 95% CI, 0.87–4.69; $P = 0.101$). To control for guarantee-time bias in the logistic regression model assessing ORR, a landmark approach was employed wherein only the 65 patients on trial ≥9 weeks were included and only those trAEs that occurred ≤9 weeks after initiating therapy were considered (Table 3).

### Treatment-related AE subgroup analyses

The same Cox proportional hazards and logistic regression models discussed above were utilized to assess the association between clinical outcomes and the presence or absence of the following four independent trAE subcategories: treatment-related select AEs ($n = 47$), possible trAE ($n = 53$), probable trAE ($n = 32$), and a grade 2 or higher trAE ($n = 31$). The unadjusted Cox proportional hazards regression model revealed a trend toward improved PFS in patients with increasing numbers of treatment-related select AEs, whereas the adjusted model revealed a clear association (unadjusted model: HR, 0.76; 95% CI, 0.55–1.05; $P = 0.097$ and adjusted model: HR, 0.62; 95% CI, 0.40–0.96; $P = 0.034$). For OS, both the unadjusted and adjusted Cox proportional hazards regression models showed a trend toward improved OS in patients with increasing numbers of treatment-related select AEs (unadjusted model: HR, 0.77; 95% CI, 0.58–1.03; $P = 0.079$ and adjusted model: HR, 0.72; 95% CI, 0.49–1.05; $P = 0.088$). Neither the unadjusted or adjusted logistic regression models evidenced a clear association between a treatment-related select AE and improved ORR (unadjusted model: OR, 1.15; 95% CI, 0.53–2.50; $P = 0.729$ and adjusted model: OR, 1.67; 95% CI, 0.55–5.09; $P = 0.364$; Supplementary Table S5).

To assess the association between a more severe trAE, defined as grade 2 or higher, and clinical outcomes, we incorporated only the 31 trAEs that were grade 2 or higher into the Cox proportional hazards and logistic regression models. In the adjusted models, but not the unadjusted models, a trend toward improved PFS, OS, and ORR was seen in the patients that experienced a grade 2 or higher trAE compared with those that did not (all $P$ values in the adjusted models <0.10, whereas all $P$ values in unadjusted models >0.10). Probable trAEs were clearly associated with improved PFS and OS, in both unadjusted and adjusted models (all $P$ values < 0.015), whereas the association between a possible trAE and PFS/OS was less clear (all $P$ values in both unadjusted and adjusted models >0.10). A strong association between a possible or probable trAE and improved ORR was not observed in the logistic regression models.

### Association between abnormal thyroid indices and clinical outcomes

Because treatment-related hypothyroidism was the most predictive trAE in our cohort, ORR 83.3% (5/6), we investigated how

**Table 3a.** Cox proportional hazards models were created for PFS with number of trAEs included as a time-varying covariate

| Effect | HR (95% CI) | P |
|---|---|---|
| PFS (unadjusted model) | | |
| Number of trAEs (+1) | 0.81 (0.65–1.01) | 0.067 |
| PFS (adjusted model) | | |
| Number of trAEs (+1) | 0.75 (0.56–0.99) | 0.043 |
| Female | 0.80 (0.48–1.35) | 0.409 |
| Age at trial start (+1 y) | 0.97 (0.95–1.00) | 0.036 |
| Number of previous lines of therapy (+1) | 1.03 (0.86–1.24) | 0.760 |
| PD-L1 proportion score (Ref: ≤1) | | |
| 1–49 | 1.13 (0.63–2.02) | 0.686 |
| ≥50 | 1.03 (0.46–2.28) | 0.948 |
| EGFR mutations | 1.51 (0.80–2.84) | 0.204 |
| Ever smoker | 0.93 (0.48–1.80) | 0.820 |
| Squamous | 1.01 (0.52–1.95) | 0.972 |

NOTE: Listed are results from both unadjusted and adjusted models. HR estimates marked "(+1)" correspond to the incremental effect of a unit increase in the relevant predictor.

**Table 3b.** Cox proportional hazards models were created for OS with number of trAEs included as a time-varying covariate

| Effect | HR (95% CI) | P |
|---|---|---|
| OS (unadjusted model) | | |
| Number of trAEs (+1) | 0.77 (0.63–0.94) | 0.011 |
| OS (adjusted model) | | |
| Number of trAEs (+1) | 0.75 (0.58–0.96) | 0.021 |
| Female | 0.92 (0.56–1.51) | 0.738 |
| Age at trial start (+1 y) | 0.99 (0.96–1.01) | 0.285 |
| Number of previous lines of therapy (+1) | 1.08 (0.91–1.28) | 0.363 |
| PD-L1 proportion score (Ref: ≤1) | | |
| 1–49 | 1.18 (0.67–2.06) | 0.566 |
| ≥50 | 1.38 (0.63–3.02) | 0.420 |
| EGFR mutations | 0.74 (0.40–1.38) | 0.345 |
| Ever smoker | 0.72 (0.39–1.34) | 0.302 |
| Squamous | 0.98 (0.49–1.96) | 0.950 |

NOTE: Listed are results from both unadjusted and adjusted models. HR estimates marked "(+1)" correspond to the incremental effect of a unit increase in the relevant predictor.

**Table 3c.** Logistic regression models were created for objective response with number of trAEs included as a covariate

| Effect | OR (95% CI) | P |
|---|---|---|
| ORR (unadjusted model) | | |
| Number of trAEs (+1) | 1.27 (0.81–1.99) | 0.306 |
| ORR (adjusted model) | | |
| Number of trAEs (+1) | 2.02 (0.87–4.69) | 0.101 |
| PD-L1 proportion score (Ref: ≤1) | | |
| 1–49 | 3.94 (0.67–23.10) | 0.129 |
| ≥50 | 26.64 (2.85–249.00) | 0.004 |
| EGFR mutations | 1.07 (0.18–6.47) | 0.937 |

NOTE: To control for guarantee bias, only the 65 patients on trial ≥9 weeks were included and only those trAEs that occurred ≤9 weeks after initiating therapy were considered. Listed are results from both unadjusted and adjusted models. ORRs are compared with ORs and OR estimates marked "(+1)" correspond to the incremental effect of a unit increase in the relevant predictor.

changes in thyroid indices (TSH, fT4, and T3) related to pembrolizumab activity. A total of 97.9% (95/97) of patients treated at UCLA with available clinical follow-up data had at least a baseline set of thyroid indices available for analysis; 74.7% (68/97) of patients had ≥3 sets of values. The association between abnormal thyroid indices in the former group of 95 patients and clinical outcomes was evaluated.

A total of 32.6% (31/95) of these 95 patients had at least one abnormal TSH noted on trial. Of these 31 patients, 24 had an elevated TSH, whereas 14 had a decreased TSH recorded. Seven of these patients had both an elevated and decreased TSH at independent time points. Patients with an abnormal TSH had a higher ORR, 35.5% (11/31), than those without, 14.1% (9/64; $P = 0.0296$). Patients with an elevated TSH had an ORR of 41.7% (10/24), whereas those with a decreased TSH had an ORR of 57.1% (8/14) compared with an ORR of 14.1% (10/71) and 14.8% (12/81) for patients without an elevated or decreased TSH, respectively (elevated TSH: $P = 0.0079$; decreased TSH: $P = 0.0014$). All 7 patients with evidence of variable thyroid function on trial, defined as both an elevated and decreased TSH noted at independent time points, had an objective response on trial (Supplementary Table S6A).

We then incorporated the timing of TSH abnormality identification into our analysis because the biological implications of a baseline abnormality, prior to therapy, are different than those of an abnormality that arose on trial. In this analysis, only patients without a baseline TSH abnormality were considered to have an "acquired" TSH abnormality. The ORR for patients with an acquired TSH abnormality was numerically greater than for patients with a baseline abnormality, with an ORR of 42.9% (9/21) in the former group, compared with 20% (2/10) in the latter group ($P = 0.2617$; Supplementary Table S6D).

An abnormal fT4 or T3 on trial were also both independently associated with improved ORR. Specifically, an abnormal fT4 was noted in 10 patients, with an ORR of 50% in this group compared with 17.7% (10/85) in patients without an abnormal fT4 ($P = 0.0317$). Nineteen patients had an abnormal T3, with an ORR of 47.4% (9/19) in this group compared with 14.5% (11/76) in those patients without an abnormal T3 ($P = 0.0037$). Neither of the 2 patients with an abnormal baseline fT4 had an objective response, whereas the ORR for patients with an acquired fT4 abnormality was 62.5% (5/8). Six patients had an abnormal T3 at baseline, with an ORR of 33.3% (2/6) in this group, whereas the ORR for patients with an acquired T3 abnormality was 53.9% (7/13). Eight of 16 patients with a low T3 had an objective response (ORR: 50%), whereas only 1 of 3 patients with a high T3 had an objective response (ORR: 33.3%). Similar patterns were observed for fT4—80% (4/5) patients with a low fT4 had an objective response compared with only 28.6% (2/7) of patients with an elevated fT4 (Supplementary Table S6B–S6D).

## Discussion

This single-center, retrospective analysis revealed that a trAE predicted for improved clinical outcome with pembrolizumab. To capture all possible AEs associated with pembrolizumab therapy on trial, rather than restricting our analysis to the very low number of AEs deemed by the investigator to be irAEs, we chose to evaluate all AEs considered at least possibly related to the study drug. In fact, only 6 patients in the UCLA cohort were reported to have an irAE at least possibly related to pembrolizumab, 3 with pneumonitis and 3 with rash. However, 6 patients experienced treatment-related hypothyroidism, a well-documented AE associated with immune checkpoint inhibitors, illustrating that real-time attribution of toxicity with a new drug class can be difficult (9, 11). To limit bias introduced by this more inclusive approach to trAE inclusion, we performed a retrospective analysis to identify treatment-related select AEs and found a strong association between the occurrence of a treatment-related select AE and improved clinical outcomes. Taken together, the findings from both our trAE analysis and retrospectively filtered treatment-related select AE analysis support the hypothesis that those patients responding to therapy may also disproportionately experience treatment-related side effects (11, 15, 18).

Consistent with the experience with other immune checkpoint inhibitors in NSCLC (19), almost every NSCLC patient treated on KEYNOTE-001 at UCLA experienced an AE. However, only 10% of the AEs reported were deemed by the investigator to be treatment-related. These 85 trAE events occurred in 40% of patients on trial, which was lower than the overall numbers reported for the 495 total patients enrolled on KEYNOTE-001, inclusive of the UCLA cohort, where 70.9% of patients experienced a trAE (6). This difference may have been a result of our investigators' increased level of comfort with anti–PD-1 therapy. Specifically, although nearly 100 patients were treated at our site, many sites would have more limited experience with anti–PD-1 therapy during the time that these agents were not widely available. The significantly higher level of experience with anti–PD-1 therapy at UCLA may have led to a more discerning classification of treatment association in comparison to investigators at other sites with less experience with these therapies. Because we do not have access to adverse event data for the total study population stratified by individual investigators, we cannot further evaluate this hypothesis. However, we were able assure that the lower rate of reporting at UCLA was not simply a result of a specific investigator's assessments because internal consistency was present within the UCLA cohort with respect to the rate at which AEs were deemed treatment-related.

Differences in demographics between the UCLA cohort and the total study population may be another possible explanation for the observed difference in the percentage of patients that experienced a trAE in the UCLA cohort compared with the total study population. Specifically, there was a much higher incidence of "never smoking," less heavily pretreated, EGFR mutation–positive patients that were less likely to experience a trAE on trial in the UCLA cohort. Again, data are not available regarding the rate at which these subpopulations experienced a trAE in the total study population, so the influence of their overrepresentation in our cohort cannot be definitively identified. However, to attempt to address the possible effects of a less heavily pretreated population in the UCLA cohort, we evaluated the influence of prior radiotherapy on the subsequent development of a trAE of any type in our patient cohort and found no clear association. This is despite the fact that patients in our cohort with a history of prior TRT had more overall treatment-related pulmonary toxicity compared with those without prior TRT, as previously reported (17). Finally, guarantee-time bias, the fact that patients on trial longer would be more likely to develop a trAE, may also have contributed to the lower number of patients experiencing a trAE in the UCLA cohort compared with the total study population, as the median PFS in the total study population was almost twice as long as the UCLA cohort (111 vs. 64 days).

The occurrence of a trAE was not associated with the myriad of clinicopathologic characteristics available for evaluation, emphasizing the robust nature of this observed association—one that cannot be otherwise explained by known predictors of response to PD-1 blockade. To further evaluate the independent nature of the observed association between a trAE and improved clinical outcomes, statistical models incorporating these clinicopathologic covariates were generated. The results from these models further strengthened the notion that this association was not due to confounding variables. All trAEs experienced on trial by $\geq$2 patients have been described with anti–PD-1 blockade (8, 9, 11, 20). That said, certain trAEs appeared to be more predictive of response to anti–PD-1 therapy in our patient cohort, such as hypothyroidism, where the occurrence of treatment-related hypothyroidism was associated with objective response to therapy in all but one of the patients who experienced this trAE.

Because the CTCAE v4 criteria for grading hypothyroidism do not take into account actual thyroid indices and are based upon symptoms and need for replacement therapy, we evaluated the association between an abnormal TSH, fT4, or T3 observed on trial and clinical outcomes. Abnormalities in TSH, fT4, or T3 strongly associated with improved clinical outcomes on trial, with acquired abnormalities (first occurrence after administration of pembrolizumab) more predictive than baseline abnormalities and abnormally low values more predictive than abnormally high values. The strong correlation observed between treatment-related hypothyroidism, as well as abnormal thyroid indices, and response to pembrolizumab is consistent with the findings of others (15, 21) and certainly warrants further evaluation.

In the 97 patients treated at UCLA on KEYNOTE-001 that had follow-up available, the patients that experienced a trAE or a treatment-related select AE had improved outcomes. These findings are consistent with experience with checkpoint inhibition in melanoma, where a correlation between improved clinical outcomes and AEs related to therapy has been observed (10, 13–15). For example, those metastatic melanoma patients treated with anti–CTLA-4 therapy in combination with a peptide vaccine, who experienced grade 3/4 autoimmune toxicity, had a much higher clinical response rate, 36%, compared with those with no autoimmune toxicity, 5% (12). However, data on this subject are conflicting, as others have found no association between immune checkpoint–related AEs and response to checkpoint blockade in melanoma (16). This observed association between response to pembrolizumab and treatment-related side effects has clinical relevance, as the systemic side effects experienced as a result of pembrolizumab could serve as a proxy for response to therapy, similar to rash in EGFR tyrosine kinase inhibitor therapy (22–24).

In terms of generalizability of our findings to the use of PD-1/PD-L1 blockade in NSCLC, it must be understood that by its very nature, any analysis assessing an association between AEs, as reported prospectively by investigators, and clinical data is entirely reliant on an individual investigator's practice of AE reporting. Although the standardized CTCAE criteria establish a validated framework by which AEs can and should be reported, every investigator still has complete autonomy with respect to event reporting. In our analysis, we have attempted to mitigate this limitation in a number of ways. First, our central analysis was based on an agnostic approach to trAE inclusion. This approach was chosen to mitigate AE selection bias introduced by retrospective filtering of AEs for those considered most likely to be a result of therapy. As PD-1/PD-L1 blockade is a relatively new therapeutic approach in NSCLC, it seems premature to assert that we know precisely which AEs are a result of therapy and which are not. Second, subanalyses were performed to address potential sources of bias within each investigator's AE reporting pattern. However, these subanalyses are exploratory in nature, and due to multiplicity issues, the results should be interpreted with caution. That said, the association between improved clinical outcomes and a trAE remained robust when stratifying the inclusion of an AE by the degree of attribution (possible, probable) to pembrolizumab and also when restricting the analysis to more severe trAEs, defined as grade 2 or higher. Note that $\geq$95% of patients experienced an AE of any relation to pembrolizumab or an AE considered unlikely related to pembrolizumab, so stratifying patients by the presence or absence of such an AE and correlating to outcomes does not yield meaningful information. Although the benefits of an agnostic approach to trAE inclusion are clear, we did perform retrospective filtering of AEs reported on trial to include only the 47 events considered treatment-related select AEs (15). In so doing, we attempted to mitigate the bias of including too broad a group of trAEs in our analysis, inclusive of AEs that may not be related to pembrolizumab mechanistically, and the results of this analysis largely corroborated the central findings of our agnostic approach. Cox proportional hazards and logistic regression models were generated to minimize guarantee-time bias, wherein the number of trAEs was modeled as a time-varying covariate in the Cox proportional hazards regression model and a landmark approach was employed in the logistic regression models because trAE status was unknown at trial initiation. The results of these models clearly support the observed association between the occurrence of a trAE and improved clinical outcomes in our patient cohort and suggest this association is largely independent of guarantee-time bias. However, the seven factors included in the multivariable models were nearly all nonsignificant in their predictive nature for benefit with pembrolizumab in our patient cohort, likely due to the limited number of patients available for analysis and associated study events. Finally, because some CTCAE v4 criteria are based upon symptomatology or medical interventions required rather than more objective endpoints, we performed a subanalysis of laboratory values for the most predictive trAE in our cohort, hypothyroidism, and found compelling preliminary data that the development of thyroid dysfunction on trial was highly predictive of treatment efficacy. Ultimately, it will take many more similar analyses in NSCLC to accurately characterize the type of adverse events mechanistically associated with PD-1 blockade and ascertain the generalizability of our findings, but we strongly believe that the analysis presented herein provides a reasonable starting place.

Although our retrospective analysis has the weakness of being conducted at a single center representing only approximately 20% of patients on trial, the strength is that a limited number of investigators assessed if an event was an AE and was treatment-related. Given the clear internal consistency of our data set, we employed an agnostic approach with respect to treatment-related AE inclusion, to assure all side effects of therapy were considered. The central conclusion of our analysis, namely that AEs believed to be associated with PD-1 blockade are associated with improved patient outcomes, is further supported by the retrospective treatment-related select AE analysis included. Future work will focus

on evaluating the association between trAEs and response to anti–PD-1 therapy in a larger patient population and with other NSCLC immunotherapeutics.

## Disclosure of Potential Conflicts of Interest

A. Lisberg is a consultant/advisory board member for AstraZeneca. J.W. Goldman reports receiving commercial research funding from Bristol-Myers Squibb, Merck Pharmaceuticals, and Roche/Genentech; has honoraria from the Speakers Bureau of Merck Pharmaceuticals; and is a consultant/advisory board member for Bristol-Myers Squibb. M. Mendenhall has honoraria from the Speakers Bureau of Merck Pharmaceuticals. E.B. Garon reports receiving commercial research support from AstraZeneca, Genentech, Bristol-Myers Squibb, and Merck. No potential conflicts of interest were disclosed by the other authors.

## Authors' Contributions

**Conception and design:** A. Lisberg, D.A. Tucker, E.B. Garon
**Development of methodology:** A. Lisberg, D.A. Tucker, N. Shaverdian, E.B. Garon
**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** A. Lisberg, D.A. Tucker, J.W. Goldman, J. Carroll, A. Hardy, C. Adame, C. Wells, J. McKenzie, B. Ledezma, M. Mendenhall, P. Abarca, K. Bornazyan, N. Moghadam, D. Nameth, C. Marx, J. Madrigal, N. Shaverdian, E.B. Garon
**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** A. Lisberg, D.A. Tucker, J.W. Goldman, B. Wolf, D. Nameth, S. Vangala, N. Shaverdian, D. Elashoff, E.B. Garon
**Writing, review, and/or revision of the manuscript:** A. Lisberg, D.A. Tucker, J.W. Goldman, B. Wolf, J. Carroll, M. Mendenhall, K. Bornazyan, S. Vangala, N. Shaverdian, D. Elashoff, E.B. Garon
**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** A. Lisberg, B. Wolf, P. Linares, C. Wells, J. McKenzie, P. Abarca, K. Bornazyan, N. Chong, C. Marx, E.B. Garon
**Study supervision:** A. Lisberg, E.B. Garon

## Acknowledgments

E.B. Garon was in part supported by the One Ball Matt Memorial Golf Tournament and R01 CA208403.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 2, 2017; revised July 19, 2017; accepted January 19, 2018; published OnlineFirst January 30, 2018.

## References

1. Golomb HM, Jacobs A, Fefer A, Ozer H, Thompson J, Portlock C, et al. Alpha-2 interferon therapy of hairy-cell leukemia: a multicenter study of 64 patients. J Clin Oncol 1986;4:900–5.
2. Quesada J, Hersh E, Manning J, Reuben J, Keating M, Schnipper E, et al. Treatment of hairy cell leukemia with recombinant alpha-interferon. Blood 1986;68:493–7.
3. Quesada JR, Reuben J, Manning JT, Hersh EM, Gutterman JU. Alpha interferon for induction of remission in hairy-cell leukemia. N Engl J Med 1984;310:15–8.
4. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. N Engl J Med 2015;373:1627–39.
5. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non–small-cell lung cancer. N Engl J Med 2015;373:123–35.
6. Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non–small-cell lung cancer. N Engl J Med 2015;372:2018–28.
7. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1–positive non–small-cell lung cancer. N Engl J Med 2016;375:1823–33.
8. Champiat S, Lambotte O, Barreau E, Belkhir R, Berdelou A, Carbonnel F, et al. Management of immune checkpoint blockade dysimmune toxicities: a collaborative position paper. Ann Oncol 2016;27:559–74.
9. Naidoo J, Page DB, Li BT, Connell LC, Schindler K, Lacouture ME, et al. Toxicities of the anti-PD-1 and anti-PD-L1 immune checkpoint antibodies. Ann Oncol 2015;26:2375–91.
10. Weber JS, Kahler KC, Hauschild A. Management of immune-related adverse events and kinetics of response with ipilimumab. J Clin Oncol 2012;30:2691–7.
11. Spain L, Diem S, Larkin J. Management of toxicities of immune checkpoint inhibitors. Cancer Treat Rev 2016;44:51–60.
12. Attia P, Phan GQ, Maker AV, Robinson MR, Quezado MM, Yang JC, et al. Autoimmunity correlates with tumor regression in patients with metastatic melanoma treated with anti–cytotoxic T-Lymphocyte antigen-4. J Clin Oncol 2005;23:6043–53.
13. Downey SG, Klapper JA, Smith FO, Yang JC, Sherry RM, Royal RE, et al. Prognostic factors related to clinical response in patients with metastatic melanoma treated by CTL-associated antigen-4 blockade. Clin Cancer Res 2007;13:6681–8.
14. Phan GQ, Yang JC, Sherry RM, Hwu P, Topalian SL, Schwartzentruber DJ, et al. Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. Proc Natl Acad Sci USA 2003;100:8372–7.
15. Weber JS, Hodi FS, Wolchok JD, Topalian SL, Schadendorf D, Larkin J, et al. Safety profile of nivolumab monotherapy: a pooled analysis of patients with advanced melanoma. J Clin Oncol 2017 ;35:785–92.
16. Grimaldi A, Simeone E, Festino L, Giannarelli D, Palla M, Caracò C, et al. Correlation between immune-related adverse events and response to pembrolizumab in advanced melanoma patients. J Immunother Cancer 2015;3:P186.
17. Shaverdian N, Lisberg AE, Bornazyan K, Veruttipong D, Goldman JW, Formenti SC, et al. Previous radiotherapy and the clinical activity and toxicity of pembrolizumab in the treatment of non-small-cell lung cancer: a secondary analysis of the KEYNOTE-001 phase 1 trial. Lancet Oncol 2017;18:895–903.
18. Freeman-Keller M, Kim Y, Cronin H, Richards A, Gibney G, Weber JS. Nivolumab in resected and unresectable metastatic melanoma: characteristics of immune-related adverse events and association with outcomes. Clin Cancer Res 2016;22:886–94.
19. Fehrenbacher L, Spira A, Ballinger M, Kowanetz M, Vansteenkiste J, Mazieres J, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. Lancet 2016;387:1837–46.
20. Bristol-Myers Squibb. Opdivo: immune-mediated adverse reactions management guide. Available from: http://www.opdivoyervoyhcp.com/servlet/servlet.FileDownload?file=00Pi000000SRPWfEAP.
21. Leighl N, Gandhi L, Hellman M, Horn L, Ahn M-J, Garon E, et al. Pembrolizumab for NSCLC: immune-mediated adverse events and corticosteroid use. In: Proceeding of the 16th World Conference on Lung Cancer. Denver, Colorado, 2015.
22. Lee Y, Shim HS, Park MS, Kim J-H, Ha S-J, Kim SH, et al. Gene copy number and skin rash as predictive markers for EGFR tyrosine kinase inhibitors in patients with advanced squamous cell lung carcinoma. Clin Cancer Res 2012;18:1760–8.
23. Liu H-b, Wu Y, Lv T-f, Yao Y-w, Xiao Y-y, Yuan D-m, et al. Skin rash could predict the response to EGFR tyrosine kinase inhibitor and the prognosis for patients with non-small cell lung cancer: a systematic review and meta-analysis. PLoS ONE 2013;8:e55128.
24. Mohamed MK, Ramalingam S, Lin Y, Gooding W, Belani CP. Skin rash and good performance status predict improved survival with gefitinib in patients with advanced non-small cell lung cancer. Ann Oncol 2005;16:780–5.