**Validation of a clinical blood-based decision aid to guide immunotherapy treatment in patients with non-small cell lung cancer.**

Mirte Muller, MD1, Roland Hoogendoorn, MD1, Ruben Moritz2, Vincent van der Noort, PhD3, Mirthe Lanfermeijer2, Catharina M. Korse, PhD2, Daan van den Broek, PhD2, Jelle J. ten Hoeve4, Paul Baas, PhD1, Huub H. van Rossum, PhD2 , Michel M. van den Heuvel, PhD1,5\*

1Department of Thoracic Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands, 2Department of Laboratory medicine, The Netherlands Cancer Institute, Amsterdam, The Netherlands, 3Department of biometrics, The Netherlands Cancer Institute, Amsterdam, The Netherlands, 4Division of molecular carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands, 5 Department of respiratory diseases, Radboud Medical Center, Nijmegen, The Netherlands.

Supplemental data

**Contents:**

Methods – supplemental

Figure S1 – Schematic view

Figure S2 – Cut-off check training set

Figure S3 – Week 2-20 for the training cohort

Figure S4 – Week 2-20 for the validation cohort

Figure S5 – Multiple biomarker testing (Also: table S1)

Figure S6 – Tumor marker dynamics and interpretation in representative examples

Table S1 – Considerations for the development of a marker test

Table S2 – Multiple biomarker testing in training set (also: Figure S4)

Table S3 – Table of patient characteristics for week 6

Table S4 – Number of available test per individual marker in training and validation set

Table S5 – False-positives for test at week 6.

Table S6 – Sub analysis Squamous Cell Carcinoma and SCC marker

Table S7 – Patient characteristics of the first line pembrolizumab cohort

Table S8 – Analyses first line pembrolizumab cohort

**Methods**

**Study population**

All patients received their treatment in a variety of settings, such as routine care, expended access, compassionate use programs or clinical trials. Patient criteria for the start of immunotherapy are more or less the same and are mentioned below.

**Nivolumab**

**Compassionate use program**

From august 2015 until the spring of 2016 nivolumab was provided in a compassionate use program (NCT02475382) by Bristol-Myers Squibb. These patients are already described in the publication from Schouten et al and more information about the eligibility criteria and protocol can be found there as well. (1)

**Standard of care**

In the spring of 2016 nivolumab became available as routine clinical care for patients with both advanced squamous non-small cell lung cancer or non-squamous. The inclusion criteria were less strict compared to the compassionate use program, however, patients still had received at least one previous line of anticancer treatment, results of blood tests consistent with adequate organ functions, and a good clinical performance. These patients are already described in the publication from Schouten et al (1).

**Pembrolizumab**

**Second line**

19 patients who were treated with pembrolizumab in second line, were participating in a trial with pembrolizumab (NCT02492568). This was a randomized phase II, 2-arm study of pembrolizumab after high dose radiation (SBRT) versus pembrolizumab alone in patients with advanced non-small cell lung cancer). Only patients who did not receive SBRT were included in the current analysis. The other patients who received pembrolizumab in the second line of treatment, were treated according to standard of care. This meant they all had a PD-L1 score of at least 1% at baseline.

**First line**

Patients who were treated with pembrolizumab in the first line(Table S6 and S7), were treated according to standard of care. This meant they all had a PD-L1 score of >50% at baseline.

Furthermore, there were 8 patients (2.1%) included in the second line analysis of immunotherapy, but are called first line in table 1. These patients did receive a first line of treatment, however, too short to expect a response from that therapy. Reasons were for example: patient did not want to continue chemotherapy soon after the first gift and therefore received the nivolumab soon after the chemotherapy.

**Statistical analysis**

**Sample size rationale**

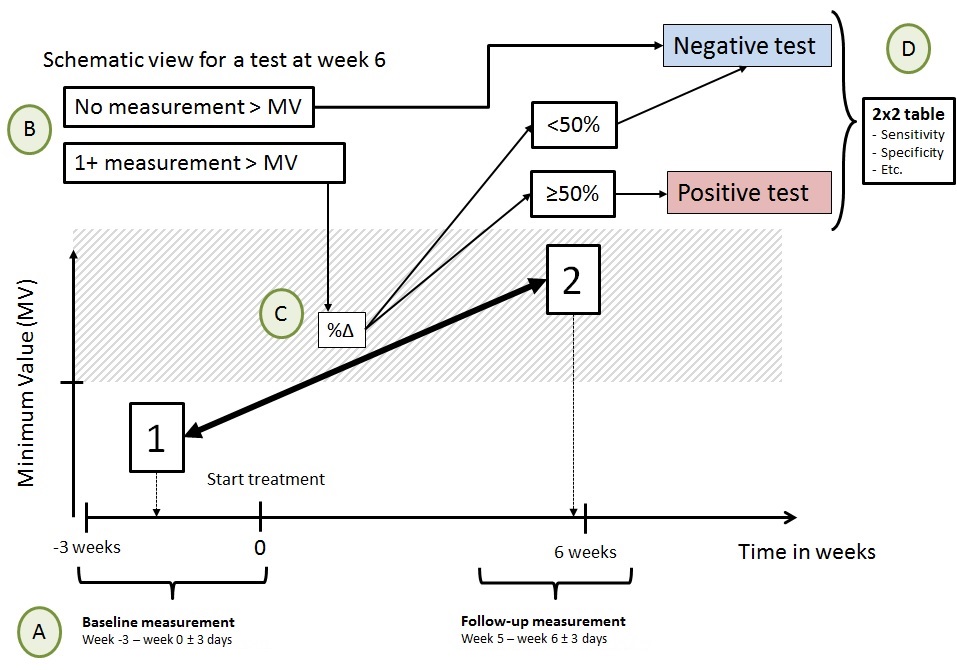
At the start of the study, no actual sample size calculation was performed. The rationale for that was that the measurement of the markers was a standard procedure in the treatment of the patients. Therefore, the design of our study was to include all patients who were eligible. In fact, inclusion in this cohort is still ongoing.

However, of course we needed to decide how many patients would go into the training and validation cohorts used in the analysis presented here. As described in the main text, in January 2017 all patients who had been treated with nivolumab at that time were randomly assigned to the training or validation cohort in a 2:1 ratio (figure 1), while subsequent eligible patient were assigned to the validation cohort.

This was based on the following considerations.

* We aimed for a 1:1 ratio between training and validation patients for the simple reason that making the training cohort too small would lead to over-fitting and making the validation cohort too small would diminish clinical usefulness of our results. Similarly we want both cohorts to be representative of the underlying population hence not artificially enrich either cohort with a higher number of responders.
* When in the Netherlands nivolumab became available for routine clinical care for second line treatment of NSCLC, in Q2 2016, the manuscript of Herbst et al was published about first-line treatment with pembrolizumab in NSCLC patients with a PD-L1 score >50% (2). Therefore, in March 2017, less than a year following the introduction of nivolumab in the Netherlands, soon thereafter the food and drug administration (FDA) approved pembrolizumab for NSCLC patients with a PD-L1 >50%. The FDA approval for the first line treatment in combination with chemotherapy for patients with a PD-L1 <50% followed soon. Based on the considerations, that patients, treated later in time during this period, were different and treated differently, we decided in January 2017 to randomly assign patients in a 2:1 ratio to the training and validation cohort thereby correcting for treatment and patient variation in time and reducing the number of patient needed for the validation.
* It was decided on forehand that only tests with at least 97.5% specificity in the training set would be validated in the validation set. Assuming that the validation set would show a specificity of 97.5% for the given test, we decided that the lower bound of the confidence interval around this number needed to be at least 90% for the test to be useful in clinical practice, for the prediction of non-response. This required a number of at least 50 responders in the validation cohort.
* The percentage of roughly 30% responders in the second line setting (1, 3, 4), these considerations naturally lead to the numbers of 180 patients per cohort (training and validation) used here.

**Figure S1 – Schematic view of the application.**

****

**Schematic view of the ReMarker application.**

**Figure S1: A:** The application checks if a patient has measurements (in figure referred to as 1 and 2) within the given timeframes. If there is more than one measurement in a given baseline or follow-up timeframe, the last measurement will be used. If measurements within one of the two timeframes are missing, the results are obviously defined as non-conclusive (and therefore will not be add to the 2x2 cross table).  **B:** The program checks if at least one of these measurement is above the minimum value (MV). The minimum value is a threshold used to avoid background noise from non-relevant increases of a marker.If both values are below this threshold, the test will be defined as negative. **C:** If abovementioned criteria are fulfilled, than the difference in percentage (%Δ = (measurement 2 – measurement1) / measurement1 x 100%) will be calculated**.** In our study, and in alignment with the results from the ReMarker analysis , the test was deemed to be positive when the marker in a patient had increased with 50%. **D:** The outcome of the test of a patient is used in a 2x2 cross table and compared to clinical data.

**Description of ReMarker:** The ReMarker application (5) is designed to study serial markers in multiple clinical settings and diseases. The application is able to relate titer changes to clinical outcome (non-response vs. response). Furthermore, the application allows us to change the settings in order to accommodate for treatment requirements. For example, treatment schedules can differ depending on the drugs (e.g. 2, 3, 4, or 6-weekly IV administrations or oral drugs). In the current study the drugs nivolumab and pembrolizumab were given at 2- and 3-weekly intervals respectively and therefore it was convenient to set the follow-up measurement on week 5-6 and week 5-7 respectively. The algorithm also allows adaptation to the relevant clinical indication. For instance, a given clinical follow-up period, used to get an optimal insight in the efficacy of the treatment, differs depending on the disease, the stage of disease, and / or the line of treatment. The algorithm also allows alteration of the definition of responders and non-responders (e.g. a patient with stable disease can be placed either in the non-response or response group). More technical details can be found in the paper of Moritz et al (5).

|  |  |
| --- | --- |
| Minimum value U/ml 🡪    **A. CA125** |  |
| Minimum value μg/L 🡪  **B. CEA** | Minimum value μg/L 🡪  **C. CYFRA21.1** |
| Minimum value ug/L 🡪  **D. NSE** | Minimum value μg/L 🡪    **E. SCC** |
| **Figure S2 – The optimal minimum value at week 6 in the training cohort**  μg/L: microgram per liter; U/ml: Units per milliliter. | |

|  |  |
| --- | --- |
| Weeks 🡪  **A. CA125 (65 U/ml)** |  |
| Weeks 🡪  **B. CEA (6 μg/L)** | Weeks 🡪  **C. Cyfra21.1 (4 μg/L)** |
| Weeks 🡪  **D. NSE (20 μg/L)** | Weeks 🡪  **E. SCC (3.5 μg/L)** |
|  |  |
| **Figure S3 – Characteristics of the serum tumor marker cut-off values for week 2 – 20 in the training cohort, shown as specificity and sensitivity.** The horizontal as indicates the tests done every other week. Every time point displayed is that week and the week before (i.e. the time period for week 2 is week 1-2). If there was more than one measurement in this time period, the latest measurement was taken. The two, straight lines indicate 20% and 95% respectively and are choosen for improved visibility.*μg/L: microgram per liter; U/ml: Units per milliliter.* | |

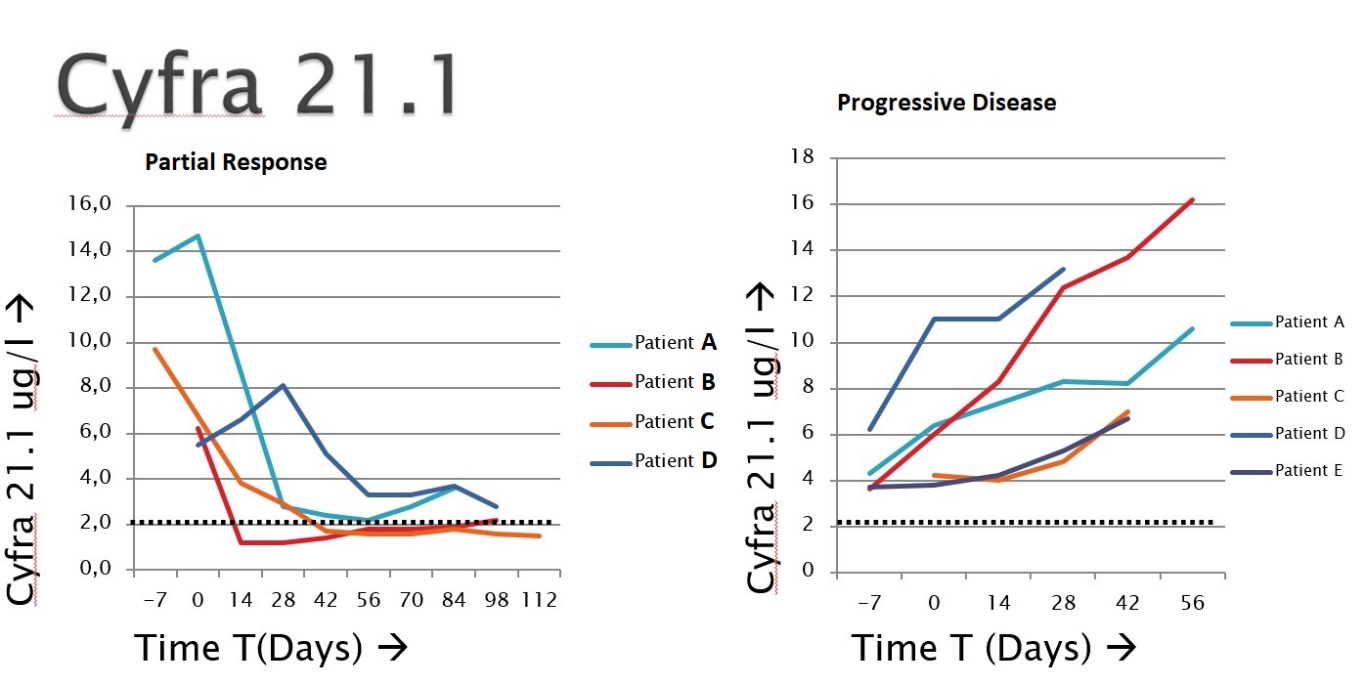
|  |  |
| --- | --- |
|  |  |
| Weeks 🡪  **A. CA125 (65 U/ml)** | Weeks 🡪  **B. SCC (3.5 μg/L)** |
| Weeks 🡪  **C. Cyfra 21·1 (8 μg/L), CEA (12 μg/L)** | Weeks 🡪  **D. Cyfra 21·1 (8 μg/L), CEA (12 μg/L), NSE (40 μg/L)** |
|  |  |
| **Figure S4 – Results of the test characteristics of tumor markers in the validation set, shown as sensitivity and specificity per week.**  The combination of markers were considered positive if at least one of the tumor markers had a positive test result. The horizontal as indicates the tests done every other week. Every time point displayed is that week and the week before (i.e. the time period for week 2 is week 1-2). If there was more than one measurement in this time period, the latest measurement was taken. The two, straight lines indicate 20% and 95% respectively and are choosen for improved visibility. *μg/L: microgram per liter; U/ml: Units per milliliter.* | |

**Figure D: NSE**

**Figure C: CYFRA**

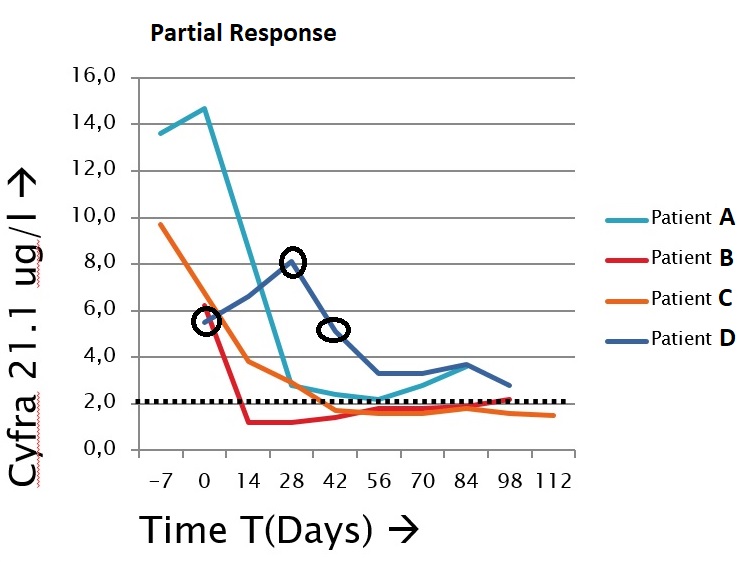
|  |
| --- |
|  |
| **Figure S4 – Multiple biomarker testing for week 6 in the training set**  The tests with a specificity of >97,5% and a sensitivity of >30% are represented. For more details, see table S1. Used cut-off values per individual marker were:CA125 65U/ml, CEA 6μg/L, Cyfra21.1 4μg/L, NSE 20μg/L and SCC 3.5μg/L. A test was considered positive when at least 1 of the individual tests was positive. *μg/L: microgram per liter; U/ml: Units per milliliter.* |

**Figure E: SCC**



A

B



**Figure S6: Tumor marker dynamics and interpretation in representative examples**

A. Tumor marker dynamics of 9 patients with partial response (left) and progressive disease (right). All these patients have at least one measurement above a cut off set by us of 4.0 ug/L. B: Interpretation of an increase/spike in our data set, an example. In our study, we designed a test which compares baseline to a certain time point during treatment. In this case, for patient D, the baseline value (5.2 ug/L) is compared to the value after 6 weeks (5.0 ug/L). Since there is no increase of >50%, this test would give a negative result and therefore be interpreted as non-progressive. It would ‘ignore’ the spike. When the values for week 4 are used (5.2 ug/L and 8.0 ug.L), the test would show up as a positive result and therefore be interpreted as progressive, or: false-positive.

**Table S1 – Considerations for the development of a marker test.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Considerations** | |  |  |
|  | |  | **Single marker test** |
| **1** | **Easy to use in clinical practice** | |  |
| **2** | **>50% increase** | | * >50% increase compared to baseline * Based on earlier BReC plot analysis (5) * Considered easy-to-calculate |
| **3** | **Minimum value** | | * The minimum value is used *instead* of the reference value of a marker. Reference values in the clinical chemistry are based on the fact that 95% of all test results lay below the reference value, measured in the healthy population (6). However, the aim is to identify non-responders in a group of lung cancer patients. Therefore, we introduced a new value, the ***minimum value***. * The role of the minimum value criterion is applied to exclude patients with small biomarker increases at low concentrations that results in large relative increases thereby reducing the effect of (pre-) analytical and biological “background noise” * At least one of the measurements should be above the minimum value (as described in figure S1) |
| **4** | **Optimal minimum value** | | * The optimal minimum value per marker was determined by calculating the ***specificity*** and ***sensitivity*** for predicting the clinical endpoint per minimum value (figure S2) |
| **5** | **Specificity > 97,5%** | | * The test was developed as an early treatment decision tool and should be very accurate in detecting non-responsiveness (to safely discontinue treatment) * Minimum values yielding a specificity of ≥97,5% in the training set per individual markers were considered a good cut-off, in order to increase the likelihood to achieve a specificity of >95% in the validation set |
| **6** | **Sensitivity >20%** | | * The test should have added value over the current standard, i.e. decisions based on radiological and clinical assessment. * Observation in the training set is that about 20% of the patients discontinue treatment within 6 weeks based on radiological assessment and/or experienced clinical deterioration. * Therefore a sensitivity >20% was considered a good cut-off. * A combination of tumor markers was assumed to increase in sensitivity, therefore for the single biomarkers a sensitivity less than 20% was accepted in the training cohort |
|  |  | | **A test with a combination of markers** |
| **7** | **Combination of single marker tests** | | * Two single marker tests with the previous described characteristics are used of a combination test |
| **8** | **At least one test is positive** | | * The combination of markers were considered positive if at least one of the tumor markers had a positive test result, defined by an increase of the marker concentration according to abovementioned criteria * When at least one of the markers is positive, the sensitivity of the test is likely to increase. |
| **9** | **Two times the minimum value** | | * Considered easy to calculate * The combination of markers may lead to a decrease in specificity, since all patients with a false-positive tests are added together in the combination test, leading to more false-positive results. |

**Table S2:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number of  markers** | **Tests** | **Sensitivity (95% CI)** | | | **Specificity (95% CI)** | | | **Positive Predicted Value(95% CI)** | | | **Negative Predicted Value (95% CI)** | | |
| 3 | Cyfra21.1, CEA, NSE | 38,4% | 28,1% | 49,5% | 100,0% | 92,8% | 100,0% | 100,0% | 89,4% | 100,0% | 48,0% | 38,0% | 58,2% |
| 2 | Cyfra21.1, NSE | 32,9% | 23,1% | 44,0% | 100,0% | 92,8% | 100,0% | 100,0% | 87,7% | 100,0% | 46,2% | 36,5% | 56,2% |
| 2 | Cyfra21.1, CEA | 38,4% | 28,1% | 49,5% | 100,0% | 92,8% | 100,0% | 100,0% | 89,4% | 100,0% | 48,0% | 38,0% | 58,2% |
| 2 | CEA, NSE | 23,3% | 14,8% | 33,6% | 100,0% | 92,8% | 100,0% | 100,0% | 83,2% | 100,0% | 42,6% | 33,4% | 52,2% |
| 4 | Cyfra21.1, CEA, NSE, CA125 | 43,0% | 32,4% | 54,2% | 98,0% | 89,2% | 100,0% | 97,4% | 86,2% | 99,9% | 49,5% | 39,2% | 59,8% |
| 3 | Cyfra21.1, NSE, SCC | 36,5% | 26,3% | 47,6% | 98,0% | 89,2% | 100,0% | 96,9% | 83,8% | 99,9% | 47,1% | 37,1% | 57,2% |
| 3 | Cyfra21.1, NSE, CA125 | 38,8% | 28,4% | 50,0% | 98,0% | 89,2% | 100,0% | 97,1% | 84,7% | 99,9% | 48,0% | 37,9% | 58,2% |
| 3 | Cyfra21.1, CEA, SCC | 41,9% | 31,3% | 53,0% | 98,0% | 89,2% | 100,0% | 97,3% | 85,8% | 99,9% | 49,0% | 38,7% | 59,3% |
| 3 | Cyfra21.1, CEA, CA125 | 43,0% | 32,4% | 54,2% | 98,0% | 89,2% | 100,0% | 97,4% | 86,2% | 99,9% | 49,5% | 39,2% | 59,8% |
| 3 | CEA, NSE, CA125 | 32,6% | 22,8% | 43,5% | 98,0% | 89,2% | 100,0% | 96,6% | 82,2% | 99,9% | 45,3% | 35,6% | 55,3% |
| 3 | CEA, NSE, SCC | 31,4% | 21,8% | 42,3% | 98,0% | 89,2% | 100,0% | 96,4% | 81,7% | 99,9% | 44,9% | 35,2% | 54,8% |
| 2 | CEA, CA125 | 31,4% | 21,8% | 42,3% | 98,0% | 89,2% | 100,0% | 96,4% | 81,7% | 99,9% | 44,9% | 35,2% | 54,8% |
| 2 | Cyfra21.1, SCC | 35,3% | 25,2% | 46,4% | 98,0% | 89,2% | 100,0% | 96,8% | 83,3% | 99,9% | 46,6% | 36,7% | 56,7% |
| 2 | Cyfra21.1, CA125 | 37,7% | 27,4% | 48,8% | 98,0% | 89,2% | 100,0% | 97,0% | 84,2% | 99,9% | 47,5% | 37,5% | 57,7% |
| 2 | NSE, CA125 | 28,2% | 19,0% | 39,0% | 98,0% | 89,2% | 100,0% | 96,0% | 79,7% | 99,9% | 44,0% | 34,5% | 53,9% |
| 2 | CEA, SCC | 27,9% | 18,8% | 38,6% | 98,0% | 89,2% | 100,0% | 96,0% | 79,7% | 99,9% | 43,6% | 34,2% | 53,4% |
| 2 | NSE, SCC | 22,6% | 14,2% | 33,1% | 97,9% | 88,9% | 100,0% | 95,0% | 75,1% | 99,9% | 42,0% | 32,7% | 51,7% |
| 5 | Cyfra21.1, CEA, NSE, CA125, SCC | **46,5%** | 35,7% | 57,6% | **95,9%** | 86,0% | 99,5% | **95,2%** | 83,8% | 99,4% | **50,5%** | 40,0% | 61,1% |
| 4 | Cyfra21.1, CEA, CA125, SCC | 46,5% | 35,7% | 57,6% | 95,9% | 86,0% | 99,5% | 95,2% | 83,8% | 99,4% | 50,5% | 40,0% | 61,1% |
| 3 | Cyfra21.1, CA125, SCC | 41,2% | 30,6% | 52,4% | 95,9% | 86,0% | 99,5% | 94,6% | 81,8% | 99,3% | 48,5% | 38,2% | 58,8% |
| 3 | CEA, CA125, SCC | 37,2% | 27,0% | 48,3% | 95,9% | 86,0% | 99,5% | 94,1% | 80,3% | 99,3% | 46,5% | 36,6% | 56,7% |
| 3 | NSE, CA125, SCC | 34,1% | 24,2% | 45,2% | 95,9% | 86,0% | 99,5% | 93,6% | 78,6% | 99,2% | 45,6% | 35,8% | 55,7% |
| 2 | CA125, SCC | 27,1% | 18,0% | 37,8% | 95,9% | 86,0% | 99,5% | 92,0% | 74,0% | 99,0% | 43,1% | 33,7% | 53,0% |

**Table S2 – The results of the tests with multiple markers in the training set (also see figure S4)** Used cut-off values per individual marker were: CA125 65U/ml, CEA 6μg/L, Cyfra21.1 4μg/L, NSE 20μg/L and SCC 3.5μg/L. A test was considered positive when at least 1 of the individual tests was positive. μg/L: microgram per liter; U/ml: Units per milliliter.

**Table S3 - Patient characteristics for patients with a test at week 6**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **FULL cohort** | **TRAINING** |  |  | **VALIDATION** |  |  |  |  |
|  | **Non-responders**  **(PD)** | **Responders**  **(PR & SD)** |  | **Non-responders**  **(PD)** | **Responders**  **(PR & SD)** |  | **Total** |  |
|  | **N=86** | **N=49** | ***P-value*** | **N=89** | **N=61** | ***P-value*** | **N=285** | ***P-value*** |
| **Patient**  Male sex – no. (%)  Age (years) – mean (SD)  Smoking – no. (%)  Pack years – mean (SD)  WHO ≥ 2– no.(%) | 50 (58.1)  63.8 (SD: 10.0)  14 (16.3)  30.1 (SD: 17.9)  11 (12.8) | 25 (51.0)  64.1 (SD: 8.0)  3 (6.1)  35.5 (SD: 19.9)  2 (4.1) | *0.423*  *0.837*  *0.053*  *0.144*  *0.091* | 47 (52.8)  63.1 (SD: 9.1)  14 (15.7)  35.2 (SD: 19.6)  6 (6.7) | 30 (49.2)  61.8 (SD: 9.1)  3 (4.9)  35.0 (SD: 17.6)  3 (4.9) | *0.662*  *0.883*  *0.108*  *0.048*  *0.649* | 152 (55.3)  63.2 (9.2)  34 (11.9)  33.7 (18.8)  22 (7.7) | *0.476*  *0.585*  *0.853*  *0.296*  *0.286* |
| **Tumor characteristics**  **Pathology**– no.(%)  Adenocarcinoma  Squamous  Other  **Mutations**– no.(%)  EGFR positive  KRAS positive  BRAF  ALK  **PD-L1** – no.(%)  Unknown  PD-L1 <1%*6*  PD-L1 >1%*6*  PD-L1 >50%*36*  Brain Metastasis – no.(%) | 58 (67.4)  21 (24.4)  7 (8.1)  1 (1.2)  24 (27.9)  3 (3.5)  0  43(50.0)  26(60.5)  17(39.5)  4(9.3)  19 (22.1) | 31 (63.3)  11 (22.4)  7 (14.3)  0  14 (28.6)  1 (2.0)  0  18 (36.7)  15 (48.4)  16 (51.6)  10 (32.3)  12 (24.5) | *0.530*  *0.441*  *0.814*  *0.521*  *-*  *0.302*  *0.013*  *0.750* | 59 (66.3)  20 (22.5)  10 (11.2)  3 (3.4)  27 (30.3)  4(4.5)  2(2.5)  40 (44.9)  29 (59.2)  20 (40.8)  9 (18.4)  14(15.7) | 36 (59.0)  15 (24.6)  10 (16.4)  2 (3.3)  22 (36.1)  1 (1.6)  1 (1.6)  23 (37.7)  14 (36.8)  24 (63.2)  20 (52.6)  10 (16.4) | *0.579*  *0.821*  *0.870*  *0.269*  *0.768*  *0.040*  *0.001*  *0.879* | 184 (64.6)  67 (23.5)  20 (7.0)  6 (2.1)  87 (30.5)  9 (3.2)  3 (1.1)  124 (43.5)  84 (52.2)  77 (42.5)  43 (26.7)  55 (19.3) | *0.740*  *0.114*  *0.060*  *0.856*  *0.063*  *0.449*  *0.000*  *0.144* |
| **Treatment**  Nivolumab  Pembrolizumab  **Line of treatment** – no.(%)  1st line  2nd line  ≥ 2nd line  Deceased after 6 months | 86 (100)  0  2 (2.3)  61 (70.9)  22 (25.6)  47 (54.7) | 49 (100)  0  0  38 (77.6)  11 (22.4)  0 | *-*  *0.486*  *-* | 72 (80.9)  17 (19.1)  2 (2.5)  64 (71.9)  23 (25.8)  31 (34.8) | 46 (75.4)  15 (24.6)  2 (3.3)  48 (78.7)  11 (18.0)  0 | *0.420*  *0.511*  *-* | 253 (88.8)  32 (11.2)  6 (2.1)  211 (74.0)  67 (23.5)  78 (20.3) | *0.000*  *0.747*  *0.007* |
| **Comorbidities**  Auto Immune Disease – no.(%) | 6 (7.0) | 0 | *0.063* | 7 (7.9) | 5 (8.2) | *0.983* | 18 (6.3) | *0.228* |

**Table S3 - Patient characteristics for patients with a test at week 6**

1 Percentages shown are based on total known PD-L1 scores.

These are the baseline characteristics of the test at 5-6 weeks.   
Abbreviations: N: Number; SD: Standard Deviation; no.: Number of patients, ECOG performance-status score: European Cooperative Oncology Group performance status score, this is a score ranging from 0 to 5, where 0 indicates no symptom, 1 indicates mild symptoms and above 1 indicates greater disability; EGFR: Epidermal Growth Factor Receptor; KRAS: Kirsten rat sarcoma viral oncogene; BRAF: v-raf murine sarcoma viral oncogene homolog B ; ALK: Anaplastic Lymphoma Kinase; PD-L1: Programmed death ligand 1.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Without cut-off value | | | With cut-off value | | | | | |
| Tumor Marker (cut-of) |  | Patients (N) | Responders | Non-responders | Patients (N) | Responders | Non-responders | Number of positive tests (>50%) | False-Positive | True positive |
| CA125 (65) | **Total** | **268** | **104** | **164** | **125** | **40** | **85** | **47** | **9** | **38** |
|  | Training | 130 | 47 | 83 | 60 | 13 | 47 | 19 | 1 | 18 |
|  | Validation | 138 | 57 | 81 | 65 | 27 | 38 | 28 | 8 | 20 |
| CEA (6) | **Total** | **284** | **110** | **174** | **170** | **58** | **112** | **35** | **1** | **34** |
|  | Training | 135 | 49 | 86 | 88 | 29 | 59 | 17 | 0 | 17 |
|  | Validation | 149 | 61 | 88 | 82 | 29 | 53 | 18 | 1 | 17 |
| CEA (12) | **Total** |  |  |  | 128 | 44 | 84 | 28 | 1 | 27 |
|  | Training |  |  |  | 65 | 22 | 43 | 13 | 0 | 13 |
|  | Validation |  |  |  | 63 | 22 | 41 | 15 | 1 | 14 |
| Cyfra 21.1 (4) | **Total** | **283** | **110** | **173** | **200** | **58** | **142** | **63** | **5** | **58** |
|  | Training | 134 | 49 | 85 | 93 | 22 | 71 | 27 | 0 | 27 |
|  | Validation | 149 | 61 | 88 | 107 | 36 | 71 | 36 | 5 | 31 |
| Cyfra 21.1 (8) | **Total** |  |  |  | 121 | 23 | 98 | 45 | 3 | 42 |
|  | Training |  |  |  | 61 | 11 | 50 | 20 | 0 | 20 |
|  | Validation |  |  |  | 60 | 12 | 48 | 25 | 3 | 22 |
| NSE (20) | **Total** | **265** | **102** | **163** | **83** | **17** | **66** | **23** | **2** | **21** |
|  | Training | 128 | 45 | 83 | 46 | 11 | 35 | 11 | 0 | 11 |
|  | Validation | 137 | 57 | 80 | 37 | 6 | 31 | 12 | 2 | 10 |
| SCC (3.5) | **Total** | **272** | **106** | **166** | **29** | **7** | **22** | **13** | **3** | **10** |
|  | Training | 131 | 48 | 83 | 16 | 3 | 13 | 9 | 1 | 8 |
|  | Validation | 141 | 58 | 83 | 13 | 4 | 9 | 4 | 2 | 2 |

**Table S4 – Number of available tests, with and without a cut-off value, for the full cohort**

**Table S4 - Number of patients included in every analysis, for a test at week 6.**

Used cut-off values per individual marker were: CA125 65U/ml, CEA 6μg/L, Cyfra21.1 4μg/L and 8μg/L , NSE 20μg/L and 40μg/L and SCC 3.5μg/L. A test was considered positive when at least 1 of the individual tests was positive. μg/L: microgram per liter; U/ml: Units per milliliter.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table S5.A. False positives in the final test (Cyfra/CEA/NSE)** | | | | | | | | | | |
| ID | Diagnosis | Treatment | Response | Cohort | CA125 | CEA | Cyfra | NSE | SCC | Explanation |
| A | Ad | Nivo | PR | Val | 424% | 225% | 218% | 89% | 7012% | Hyperthyroidism / thyroiditis |
| B | Ad | Pembro | PR | Val | 205% | - | 61% | - | - | Pseudo progression |
| C | Sq | Pembro | SD | Val | - | - | 67% | - | - | “Progressive SD” |
| D | Ad | Pembro | SD | Val | - | - | 81% | - | - | Small amount of pleural fluid |
| E | Ad | Nivo | SD | Val | 86% | - | 170% | 90% | - | Nephrodrain, relatively normal kidney function (GFR 43) |

**Table S5 – False-positives.**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table S5.B. False positives in the remaining markers (CA125 and SCC)** | | | | | | | | | | |
| ID | Diagnosis | Treatment | Response | Cohort | CA125 | CEA | Cyfra | NSE | SCC | Explanation |
| F | Ad | Nivo | SD | Val | 151% | - | - | - | - | “Progressive SD” |
| G | Sq | Nivo | SD | Val | 51% | - | - | - | - | None |
| H | Ad | Nivo | SD | Train | 59% | - | - | - | - | None, although suffering from brain infarction |
| I | Ad | Nivo | SD | Val | 111% | - | - | - | - | None |
| J | Ad | Nivo | SD | Val | 70% | - | - | - | - | Skin rash grade 2 |
| K | Ad | Pembro | PR | Val | 100% | - | - | - | - | Pseudo progression |
| L | Ad | Nivo | SD | Val | - | - | - | - | 63% | None |
| M | Sq | Nivo | SD | Train | - | - | - | - | 92% | Progressive pleural metastasis. |

**Table S5 – False positives**. **A**. For all the false-positive tests in the final test at 6 weeks after start, which was defined as a responder classified by non-responder with our test, with a possible explanation. **B**. The false positive results for the remaining markers. ID: Identification number of a patient; Response: the response after six months according RECIST criteria, with SD: Stable disease and PR: partial response ; PFS: Progression free survival in days; Ad: Adenocarcinoma; Sq: Squamous Cell Carcinoma; Nivo: nivolumab; Pembro: Pembrolizumab; Val: Validation cohort; Train: Training cohort*.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Minimum value SCC |  | Specificity | Sensitivity | Positive predicted value |
| 0.0 μg/L | All pathology | 76.4%  (67.0-83.9%) | 22.9%  (16.9-30.2%) | 60.3%  (47.2-72.2%) |
|  | Squamous | 80.8%  (60.0-92.7%) | 41.0%  (26.0-57.8%) | 76.2%  (52.5-90.9%) |
| 2.0 μg/L | All pathology | 95.3%  (88.8-98.3%) | 9.6%  (5.8-15.4%) | 76.2%  (52.4-90.9%) |
|  | Squamous | 88.5%  (68.7-97.0%) | 23.1%  (11.7-39.7%) | 75.0%  (42.8-93.3%) |
| 3.5 μg/L | All pathology | 97.2%  (91.3-99.3%) | 6%  (3-11.1%) | 76.9%  (46.0-93.8%) |
|  | Squamous | 96.2%  (78.4-99.8%) | 15.4%  (6.4-31.2%) | 85.7%  (42.0-99.2%) |

**Table S6 – Sub analysis SCC**

**Table S6 – Sub analysis SCC**

Sub analysis of the full cohort versus squamous cell only. μg/L: microgram per liter.

**Table S7 – Patient characteristics pembrolizumab first line**

|  |  |  |  |
| --- | --- | --- | --- |
| **Pembrolizumab first line** | **TRAINING** |  |  |
|  | **Non-responders**  **(PD)** | **Responders**  **(PR & SD)** |  |
|  | **N=8** | **N=23** | ***P-value*** |
| **Patient**  Male sex – no. (%)  Age (years) – mean (SD)  Smoking (never) – no. (%)  Pack years – mean (SD)  WHO ≥ 2– no.(%) | 4  57.1 (SD: 6.9)  0  35.4 (SD 27.9)  3 | 7  65.5 (SD: 9.4)  0  27.6 (SD: 13.2)  1 | *0.319*  *0.029*  *-*  *0.279*  *0.018* |
| **Tumor characteristics**  Adenocarcinoma  Squamous  Other  KRAS positive  PD-L1 >50%  Brain Metastasis – no.(%) | 6  0  2  3  8  2 | 17  2  4  16  23  1 | *0.646*  *0.155*  *-*  *0.115* |

Abbreviations: N: Number; SD: Standard Deviation; no.: Number of patients, WHO: performance-status score: World Health Organization performance status score, this is a score ranging from 0 to 5, where 0 indicates no symptom, 1 indicates mild symptoms and above 1 indicates greater disability; KRAS: Kirsten rat sarcoma viral oncogene; PD-L1: Programmed death ligand 1.

**Table S8 – Results of tumor marker test at 6 weeks (5-7 weeks).**

|  |  |  |  |
| --- | --- | --- | --- |
| **Setting** | **Sensitivity (95%-CI)** | **Specificity (95%-CI)** | **Positive Predicted Value (95%-CI)** |
| CEA 6 μg/L OR Cyfra 4 μg/L | 25,0% (4.5-64.4%) | 95,6% (76.0-99.8%) | 66,7% (12.5-98.2%) |
| CEA 6 μg/L OR Cyfra 4 μg/L OR NSE 20 μg/L | 25,0% (4.5-64.4%) | 95,6% (76.0-99.8%) | 66,7% (12.5-98.2%) |
| Cyfra 10 μg/L OR CEA 10 μg/L | 20% (3-56%) | 96% (80-100%) | 66% (15-46%) |
| Cyfra 10 μg/L OR CEA 10 μg/L OR NSE 20 μg/L | 20% (3-56%) | 96% (80-100%) | 66% (15-46%) |

A total of 62 patients were treated with pembrolizumab first line. From 27 patients there was no test: 4 patients were non-evaluable, 7 patients were lost to follow up (treatment continuation in another hospital), 20 patients missed either the baseline or follow-up measurement. Data from the remaining 31 patients was used for the analysis. The same criteria were used as in the manuscript. Patient characteristics of this cohort can be found in table S6. Results of the serum tumor marker analysis are shown in table S8.

**References**

1. Schouten RD, Muller M, de Gooijer CJ, Baas P, van den Heuvel M. Real life experience with nivolumab for the treatment of non-small cell lung carcinoma: Data from the expanded access program and routine clinical care in a tertiary cancer centre—The Netherlands Cancer Institute. Lung Cancer. 2018;126:210-6.

2. Herbst RS, Baas P, Kim DW, Felip E, Perez-Gracia JL, Han JY, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet. 2015.

3. 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. The New England journal of medicine. 2015;373(17):1627-39.

4. Brahmer J, Reckamp KL, Baas P, Crino L, Eberhardt WE, Poddubskaya E, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. The New England journal of medicine. 2015;373(2):123-35.

5. Moritz R, Muller M, Korse CM, van den Broek D, Baas P, van den Noort V, et al. Diagnostic validation and interpretation of longitudinal circulating biomarkers using a biomarker response characteristic plot. Clinica chimica acta; international journal of clinical chemistry. 2018;487:6-14.

6. Het farmacotherapeutisch kompas, Zorg instituut Nederland. https://www.farmacotherapeutischkompas.nl/farmacologie/referentiewaarden. Page seen at 18th of october 2019,