

ASSESSMENT OF THE USEFULNESS OF THE QUANTITATIVE METHODS FOR THE RESPONSE EVALUATION OF SOLID TUMORS: ANALYSIS USING FOUR CANCER TYPES

MAURICIO CANALS-LAMBARRI^{1,2,3*}, ANDREA CANALS-CIFUENTES⁴, ADOLFO BARROS-ROCCO³, PERLA BARROS-NELSON³, MAURICIO MAHAVE-CACERES⁵ AND PAMELA SALMAN-BOGHIKIAN⁵

¹Department of Medicine, Faculty of Medicine, University of Chile, Santiago, Chile; ²Department of Ecological Sciences, Faculty of Sciences, University of Chile, Santiago, Chile; ³Department of Imaging, Fundación Arturo López Pérez, Santiago, Chile; ⁴Academic Directorate, Clínica Santa María, Santiago, Chile; ⁵Department of Chemotherapy, Fundación Arturo López Pérez, Santiago, Chile

ABSTRACT

Background: The evidence provided by medical imaging techniques for the staging and follow-up is relevant in oncology. **Objectives:** The aims were (i) to compare the monitoring methods, (ii) to analyze the response variability between different tumors, and (iii) to decipher a general response curve that is independent of tumor type and drug treatment. **Methods:** We analyzed the response variability in four cancer types, looking for a general response curve independent of the tumor type and drug treatment. We compared the response of different types of lesions within each cancer type via an intra-class correlation coefficient, determining the minimum number of lesions suitable for monitoring. **Results:** The tested metrics allowed an objective evaluation of the response of solid tumors. The response was homogeneous between different cancer types. The intra-class correlation was high, allowing the monitoring of the response with a low number of lesions (2-4). The currently used metrics misrepresent the changes in the lesion volumes. Indeed, we observed non-linear overestimations of the RECIST and WHO values, which were more pronounced for the intermediate values. Additionally, the inclusion of lymphadenopathy among the target lesions produced a distortion in the evaluation of the response. **Conclusion:** The quantitative counts allowed an objective evaluation of the response of the solid tumors to therapy, showing that the response was homogeneous but variable between different types of tumors. Although the currently used metrics lead to misrepresentations of the changes in the lesion volume, they allowed setting a response pattern for tracking these lesions. (REV INVES CLIN. 2015;67:182-90)

Corresponding author: Mauricio Canals-Lambarri, mcanals@uchile.cl

Key words: RECIST. Intra-class correlation. Number of lesions. Imaging evaluation. Cancer.

Corresponding author:

*Mauricio Canals-Lambarri
Department of Ecological Sciences
Facultad de Ciencias y Departamento de Medicina
Facultad de Medicina, Universidad de Chile
Las Palmeras 3425, Ñuñoa, Santiago, Chile
E-mail: mcanals@uchile.cl

Received for publication: 30-03-2015

Accepted for publication: 19-05-2015

INTRODUCTION

The evidence that is provided via medical imaging for the diagnosis, staging, and follow-up of lesions in oncology is highly relevant¹. The interpretation of medical images should be as objective as possible because many medical/therapeutic decisions are made on their basis. In 1979, the World Health Organization (WHO) introduced criteria to determine the response of solid tumors to chemotherapy without specifying any imaging protocol (WHO, 1979)². Therefore, many groups proposed amendments that subsequently created confusion^{1,3}. In 2000, a system of measurement and monitoring for solid tumors called RECIST (Response Evaluation Criteria in Solid Tumor)³ was introduced to standardize the response criteria. This system was aimed at simplifying the WHO system that uses two cross-sectional diameters in the evaluation of a tumor mass and considers a 25% increase in the product of the diameters as a tumor progression. Instead, the RECIST metric considers only the largest diameter of the lesion. Although this appears as an oversimplification, it is aimed at correcting the arbitrary boundaries between the assessment of a response and a non-response to the therapy.

The RECIST evaluation improves the visual inspection, which is misleading because, on the one hand, the relationship between the changes in volume and diameters of a lesion depend on the size of the lesion, and on the other hand, the volume changes are not symmetrical. For example, while an increase of 2.6 mm in a 5 cm lesion is irrelevant, the same increase in a 10 mm lesion represents a doubling of the volume (a 26% increase in diameter results in twice the volume of a lesion). Conversely, a reduction in the diameter by only 20.6% represents a reduction of the tumor volume by half.

The RECIST system specifies several parameters, including the mode of image acquisition (if possible computerized tomography, CT), the minimum size of the target lesions, the measurement method, and the maximum number of target lesions. Moreover, as the criteria for progression, the RECIST system defines an increase greater than or equal to 20% in the sum of the diameters of the target lesions or in the appearance of new lesions. Subsequently, some changes were brought to the original proposal. For

example, the short diameter was to be used for lymph node lesions⁴⁻⁶.

Some discrepancies between the RECIST and WHO criteria have been documented^{1,7}. While the initial studies comparing the use of either one or two diameters for monitoring the lesions revealed a concordance of approximately 95%^{3,8-10}, subsequent studies have found that the concordance was ranging between 0.68 and 0.77¹¹. These later studies highlighted that the measurement of only one diameter does not account for the characteristic shape changes that can occur within a lesion. Additionally, the number of target lesions remains a matter of discussion, as shown by the changes between the original version (maximum of 10 lesions) and the current version of the RECIST method (maximum of five lesions). Some authors have obtained good results in monitoring colorectal cancer liver metastases by measuring only one or two lesions¹²⁻¹⁴.

Mazumdar, et al.¹⁵ proved that the number of lesions to measure depends on the correlated response that these tumors show, which can be estimated via the intra-class correlation (ICC) parameter. By performing a theoretical/empirical analysis, they showed that the number of assessed lesions should vary between three and six.

Herein, we propose two general hypotheses. First, we propose that the quantitative assessment with the RECIST and WHO methods are objective but not necessarily equivalent for the follow-up of the lesion size. Second, we anticipate that these methods only evaluate the changes in size that are related to the number of cells that were eradicated by the treatment¹⁶. Therefore, these methods might denote a homogeneous response for different tumors. Consequently, our study aimed to compare the RECIST and WHO assessment criteria and to compare the monitoring by these two methods with a proxy measurement of the tumor volume. We also aimed to analyze the response variability in four types of stage IV tumors as well as the response variability of different types of lesions within each cancer type. Lastly, we sought to identify a general response curve that is independent of tumor type and drug therapy and to determine the minimal number of lesions suitable for monitoring these tumors.

MATERIAL AND METHODS

Sample

Ninety-four consecutive non-operable patients with solid tumors were followed from 2006 to 2008 at the cancer clinic "Fundación Arturo López Pérez", Santiago, Chile. All patients were in stage IV, with a diagnosis confirmed by biopsy that included lung cancer of all histological types, with the exception of small cell lung cancer, colorectal adenocarcinoma, gastric adenocarcinoma, and renal cell carcinoma. The patients presented primary and metastatic lesions in the thorax, abdomen, and/or pelvis. Excepting in lung cancer patients, the primary lesion was not included in the analysis. All patients received only chemotherapy or targeted agents, which included pazopanib, ipilimumab, sunitinib, erlotinib, paclitaxel, or carboplatin. All patients provided informed consent. The criteria for exclusion of the focus group were the stability or progression of the lesions and the cessation of the treatment. Seventeen of these excluded patients (comparative group) were analyzed with the same methodology for comparative purposes.

Study method and metrics

The patients were studied using four-channel helical CT, with the administration of a 100 ml bolus of contrast medium. The acquisition began at 20 seconds, with the image planes being acquired every 5 mm and covering the entire thorax, abdomen, and pelvis. All patients underwent a study protocol at time of their admission in the study and were monitored monthly or bimonthly, depending on the pathology. In total, 381 lesions were followed for various durations, depending on the response of the patient, which produced 1,634 data items.

During the initial study, all of the lesions were carefully noted and described. Three to seven were chosen as the target lesions, among the most conspicuous and easy to follow tumor nodules that met the RECIST 1.0 criteria (i.e. lesions larger than 10 mm, maximum of five lesions per organ, and maximum of 10 lesions in total). These lesions were measured in two diameters (long and short), selecting the longer diameter for the total count according to the RECIST criteria, which takes into account the sum of the longest diameters of all the target lesions. The

remaining lesions were considered as non-target for the follow-up. In the subsequent follow-up studies, the procedure was repeated and the evolution of the total count was followed. At each stage of the follow-up, the patients were categorized into two categories. The first category included the patients that were stable or presented a progression of the disease determined by an increase $\geq 20\%$ in the total count increase, which could be due to the appearance of new lesions or by an evident progression of the non-target lesions. The second category included the patients who responded to the treatment. If the patient's condition remained stable or progressed (category 1), she/he was not included in the study.

For each case, a count

$$R_j = \sum_i D_i$$

was obtained, where R_j represents the RECIST count of patient j and D_i represents the largest diameter of the lesion i . To obtain a comparable response measurement, the total value (R_j) was divided by the initial value (R_{j0}), generating a relative RECIST index ($RR_j = R_j/R_{j0}$) that was independent of the initial size of the lesion. Additionally, for each lesion, the product of the transverse diameters (WHO criteria) was computed: $p_i = D_i \cdot d_i$, with D_i representing the larger and d_i the short diameter (i.e. the largest diameter orthogonal to D_i) (Fig. 1). Finally, we also obtained a proxy measurement of the lesion volume based on the ellipsoid volume formula:

$$v_i = \frac{4}{3} \pi \frac{D_i d_i^2}{8}$$

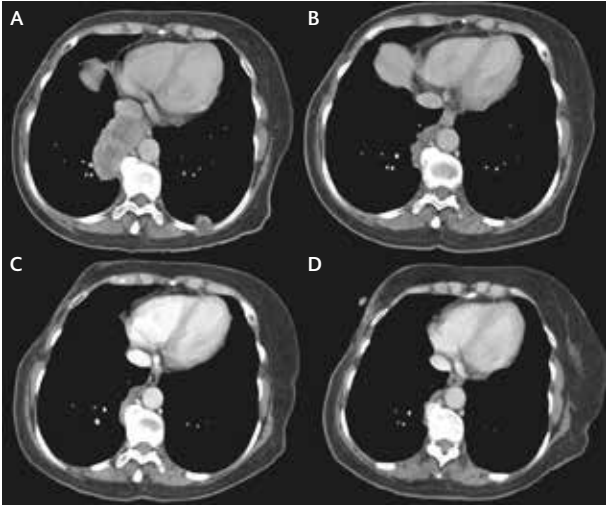
Therefore, the values of

$$R_{\text{WHO}} = \sum_i p_i$$

(i.e. the sum of the products of the diameters of the lesions) and

$$R_v = \sum_i v_i$$

Figure 1. Two target lesions in the lungs of a patient with breast cancer showing the measurement of the largest and the shortest diameters. In **A**: initial lesions, **B**: one month later, **C**: three months later, **D**: five months later.



(i.e. a proxy evaluation of the tumor volume) were obtained for the lesions and divided by the initial values to generate the relative counts RR_{WHOj} and RR_{Vj} .

Analyses

The minimal sample size of the focus group for a power of 0.9 was estimated *a priori* using the STATISTICA 7.0 software, by considering a medium size effect (i.e. root mean standardized effect, $RRME = 0.25$), a significance level of $\alpha = 0.05$, the four tumor types and four metastases types, without interaction (nested design). With these parameters, a needed sample size of 80 individuals, with 20 per cancer group, was estimated.

Overall, the response between the different types of tumors was compared using an analysis of covariance (ANCOVA), considering the RR index as the dependent variable and the exact time from the first assessment to the assessment “t” as the co-variable. The exponential regression analysis between the relative count and the time was performed to determine an overall response curve. Because not all of the patients attended their assessment exactly at monthly intervals, the data were pooled based on multiples of 30 days and a type II regression was used.

The RR , RR_{WHO} and RR_V metrics were compared using a Tukey’s mean-difference plot (Bland-Altman plot)¹⁷.

For each type of cancer, the responses between the patients and between the different types of lesions were compared using a nested ANCOVA, considering RR as the response variable.

The changes in all metrics between the focus and the comparative group were studied comparing the slopes in the exponential regressions with Student t-test for homogeneity of slopes.

From the analysis of data, we estimated the intra-class correlation among the different cancer types (ρ_{ICC}). This is a measure of the response similarity of the lesions and can be defined in general terms as¹⁵:

$$\rho_{ICC} = \frac{\text{variance (between.patients)}}{\text{Variance between.patients)} + \text{Variance (between.lesions.within.patients)}}$$

Therefore, if the variance between lesions is zero, $\rho_{ICC} = 1$. By contrast, when all the variance is a consequence of the variance between the lesions, then $\rho_{ICC} = 0$.

In our case, the response variable was RR_j and the sources of variation were the patients (P), the target lesions (L) and, considering that the lesions were followed over time (t), time was considered as a co-variable. Thus, the variance decomposition followed the following model: $RR_j = \mu + P_i + L_{l(i)} + \beta_1 t_k + \beta_0$, where the lesions were nested within the patients. Then,

$$\rho_{ICC} = \frac{\sigma^2_P}{\sigma^2_P + \sigma^2_L},$$

where σ^2 represents the variances and the subscripts P and L indicate patients and lesions, respectively.

For the intra-class correlation, the number of lesions (n^*) required to increase the variance of the RECIST count by 10% was estimated as:

$$I\sigma = \frac{V_n(R)}{V_N(R)} = \frac{N^2 [n + (n^2 - n) \rho_{ICC}]}{n^2 [N + (N^2 - N) \rho_{ICC}]},$$

where $V_n(R)$ and $V_N(R)$ are variances of the RECIST count with n and N lesions, respectively¹⁵. For the standardization, a value of $N = 10$ was used¹⁵. In other words, this analysis was designed to determine

Table 1. Distribution of all patients enrolled in this study

| | Focus group | | | Comparative group | | |
|----------------------|-------------|-----|-------|-------------------|-----|-------|
| | Women | Men | Total | Women | Men | Total |
| Lung cancer | 12 | 11 | 23 | 3 | 3 | 6 |
| Colorectal cancer | 9 | 12 | 21 | 3 | 1 | 4 |
| Gastric cancer | 10 | 19 | 29 | 3 | 2 | 5 |
| Renal cell carcinoma | 5 | 16 | 21 | 1 | 1 | 2 |
| Total | 36 | 58 | 94 | 10 | 7 | 17 |

whether measuring only n lesions among a total of N introduces a large error in the assessment of the response or not.

RESULTS

The enrolled patients included 36 women and 58 men aged between 38 and 79 years (60.3 ± 9.1 years as the mean age \pm standard deviation). The comparative group included 10 women and seven men aged between 43 and 71 years (56.5 ± 7.1 years). The patients were assigned to different therapeutic protocols according to their diagnosis of lung cancer, colorectal cancer, gastric cancer, or renal cell carcinoma (Table 1).

The RR count was similar among all the cancer types ($F_{3,12} = 1.15$; $p = 0.366$), but the lesions within the cancers displayed different behaviors ($F_{16,1632} = 4.55$; $p < 0.001$). The decrease of all metrics showed a good fit in the exponential regressions: for RR ($F_{1,28} = 81.06$; $p < 0.001$; $R^2 = 0.74$), RR_{WHO} ($F_{1,28} = 100.76$; $p < 0.001$; $R^2 = 0.78$) and RR_V ($F_{1,28} = 122.45$; $p < 0.001$; $R^2 = 0.81$) (Table 2). The decrease of all metrics, RR, RR_{WHO} and RR_V , was different between this group and the comparative group ($t_{32} = 19.5$ for RR, $t_{32} = 15.5$ for RR_{WHO} and $t_{32} = 10.6$ for RR_V ; $p < 0.001$) (Fig. 2). While in the group of patients with good response the slope was negative and significantly different from 0 for all the metrics (Table 2), in the comparative group the slope was positive (0.000588,

Figure 2. Progression in RR as function of time in the focus group (black circles) and the comparative group (white circles).

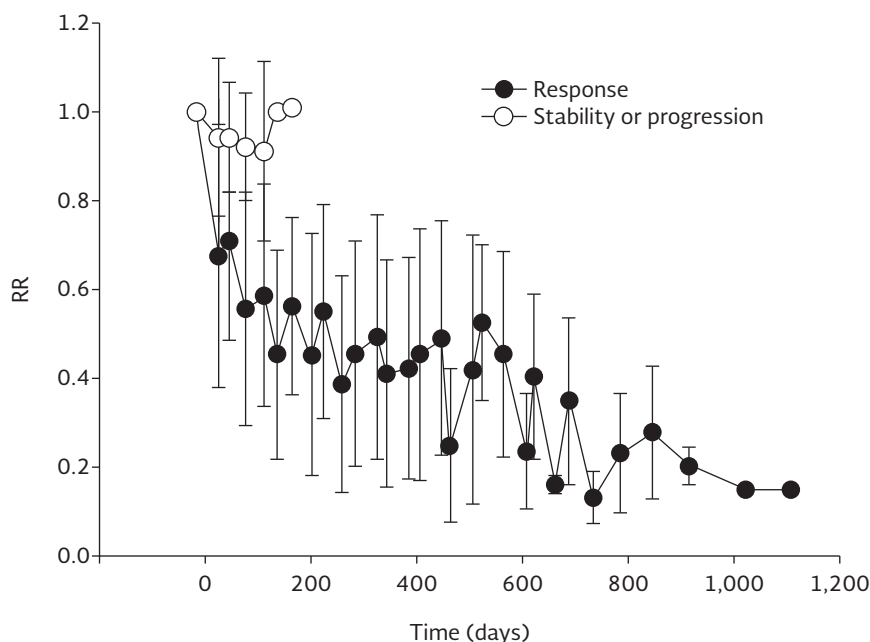


Table 2. Observed progression of the RECIST (RR), WHO (RR_{WHO}) and volumetric (RR_V) counts (means \pm standard deviation) as a function of time (months) and the predicted values (PRR , PRR_{WHO} , and PRR_V) from the exponential regression curves. $\ln(RR) = -0.315 - 0.001452t$; $\ln(RR_{WHO}) = -0.540 - 0.00221t$; $\ln(RR_V) = -0.642 - 0.00317t$. Values expressed as the proportion of the initial value (1.00 for $t = 0$)

| t (months) | RR | RR _{WHO} | RR _V | PRR | PRR _{WHO} | PRR _V |
|---------------|-----------------|-------------------|-----------------|------|--------------------|------------------|
| 2 \pm 0.28 | 0.71 \pm 0.22 | 0.54 \pm 0.26 | 0.43 \pm 0.30 | 0.67 | 0.51 | 0.44 |
| 3 \pm 0.24 | 0.56 \pm 0.26 | 0.38 \pm 0.24 | 0.29 \pm 0.23 | 0.64 | 0.47 | 0.39 |
| 5 \pm 0.29 | 0.45 \pm 0.24 | 0.29 \pm 0.20 | 0.22 \pm 0.26 | 0.59 | 0.42 | 0.33 |
| 7 \pm 0.30 | 0.45 \pm 0.27 | 0.32 \pm 0.25 | 0.24 \pm 0.26 | 0.53 | 0.36 | 0.27 |
| 9 \pm 0.34 | 0.39 \pm 0.24 | 0.24 \pm 0.17 | 0.17 \pm 0.14 | 0.49 | 0.32 | 0.22 |
| 12 \pm 0.28 | 0.42 \pm 0.25 | 0.24 \pm 0.20 | 0.15 \pm 0.15 | 0.42 | 0.25 | 0.16 |
| 15 \pm 0.17 | 0.49 \pm 0.26 | 0.30 \pm 0.19 | 0.19 \pm 0.12 | 0.38 | 0.21 | 0.12 |
| 18 \pm 0.25 | 0.46 \pm 0.23 | 0.26 \pm 0.16 | 0.15 \pm 0.11 | 0.32 | 0.16 | 0.09 |
| 21 \pm 0.41 | 0.16 \pm 0.02 | 0.06 \pm 0.13 | 0.02 \pm 0.02 | 0.28 | 0.13 | 0.06 |
| 24 \pm 0.32 | 0.13 \pm 0.06 | 0.06 \pm 0.03 | 0.03 \pm 0.02 | 0.25 | 0.11 | 0.05 |
| 30 \pm 1.04 | 0.20 \pm 0.04 | 0.08 \pm 0.01 | 0.03 \pm 0.01 | 0.19 | 0.08 | 0.03 |
| 36 \pm 1.00 | 0.15 \pm 0.05 | 0.06 \pm 0.04 | 0.02 \pm 0.05 | 0.15 | 0.05 | 0.02 |

0.000882 and 0.000904 for RR , RR_{WHO} and RR_V , respectively) and not significantly different from 0 ($p > 0.05$ in all cases).

We observed a good correlation among the different metrics: $r = 0.94$, $r = 0.87$ and $r = 0.98$ for the pairs $RR-RR_{WHO}$, $RR-RR_V$ and $RR_{WHO}-RR_V$, respectively. However, the comparisons among these metrics using the Bland-Altman plot analysis showed a poor distribution of Tukey's mean differences around the average in the three comparisons ($RR-RR_{WHO}$ shown in figure 3). This revealed an arc defect, showing that a metric based on fewer dimensions, for example a single plane, does not fully account

for a metric based on more dimensions, for example a volume.

The decrease in the size of the different types of lesions was different in the renal cell, gastric, and colorectal cancers ($F_{15, 514} = 6.94$, $p < 0.001$; $F_{12, 365} = 2.61$, $p = 0.002$ and $F_{13, 270} = 2.14$, $p = 0.012$, respectively), but was homogeneous in lung cancer ($F_{27, 347} = 1.10$; $p = 0.34$) (Table 3).

The intra-class correlations were large, varying between 0.57 and 0.87, indicating that the number of lesions (n^*) for an increment of less than 10% in the variance of the metric RR had to be between 2-4 (Table 4).

Figure 3. Arc effect in the Bland-Altman plot for the $RR-RR_{WHO}$ comparisons.

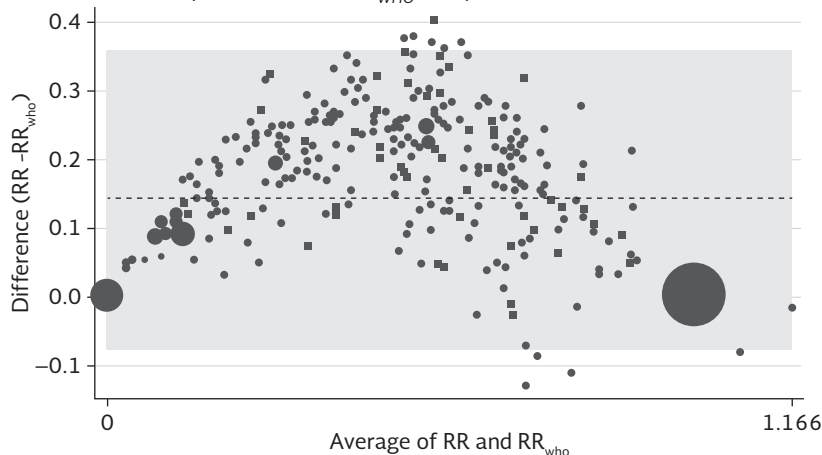


Table 3. Variation of the largest diameter (D_t) (means \pm standard deviation) of several lesions in the four types of cancer studied. The values are expressed as the proportion of the initial size (1.00 for $t = 0$)

| t (months) | Renal cell | Carcinoma | |
|--------------|-----------------|-----------------|-----------------|
| | Lung nodules | Other nodules | Adenopathy |
| 2 \pm 0.1 | 0.60 \pm 0.44 | 0.87 \pm 0.12 | 0.88 \pm 0.13 |
| 3 \pm 0.5 | 0.77 \pm 0.16 | 0.74 \pm 0.21 | 0.76 \pm 0.27 |
| 5 \pm 0.8 | 0.65 \pm 0.23 | 0.69 \pm 0.28 | 0.66 \pm 0.25 |
| 8 \pm 0.9 | 0.54 \pm 0.29 | 0.72 \pm 0.42 | 0.57 \pm 0.33 |
| 12 \pm 1.1 | 0.40 \pm 0.34 | 0.80 \pm 0.25 | 0.61 \pm 0.33 |
| 15 \pm 0.7 | 0.43 \pm 0.36 | 0.49 \pm 0.49 | 0.66 \pm 0.26 |
| 18 \pm 0.9 | 0.41 \pm 0.38 | 0.25 \pm 0.30 | 0.64 \pm 0.26 |
| 21 \pm 0.9 | 0.30 \pm 0.29 | 0.05 \pm 0.33 | 0.49 \pm 0.12 |
| | Lung cancer | | |
| | Lung nodules | Other nodules | Adenopathy |
| 2 \pm 0.1 | 0.76 \pm 0.47 | 0.86 \pm 0.28 | 0.73 \pm 0.26 |
| 3 \pm 0.5 | 0.62 \pm 0.42 | 0.84 \pm 0.22 | 0.67 \pm 0.30 |
| 5 \pm 0.8 | 0.84 \pm 0.21 | 0.71 \pm 0.32 | 0.66 \pm 0.31 |
| 8 \pm 0.9 | 0.78 \pm 0.28 | 0.58 \pm 0.22 | 0.59 \pm 0.36 |
| 12 \pm 1.1 | 0.50 \pm 0.00 | 0.61 \pm 0.11 | 0.62 \pm 0.48 |
| | Gastric cancer | | |
| | Hepatic nodules | Other nodules | Adenopathy |
| 2 \pm 0.1 | 0.80 \pm 0.32 | 0.47 \pm 0.00 | 0.77 \pm 0.17 |
| 3 \pm 0.5 | 0.77 \pm 0.22 | 0.73 \pm 0.50 | 0.63 \pm 0.26 |
| 5 \pm 0.8 | 0.66 \pm 0.33 | 0.44 \pm 0.17 | 0.51 \pm 0.38 |
| 8 \pm 0.9 | 0.48 \pm 0.36 | 0.17 \pm 0.17 | 0.39 \pm 0.32 |
| 12 \pm 1.1 | 0.27 \pm 0.18 | 0.15 \pm 0.18 | 0.37 \pm 0.34 |
| | Colon cancer | | |
| | Hepatic nodules | Lung nodules | Adenopathy |
| 2 \pm 0.1 | 0.79 \pm 0.14 | 0.67 \pm 0.25 | 0.77 \pm 0.19 |
| 3 \pm 0.5 | 0.71 \pm 0.40 | 0.79 \pm 0.12 | 0.84 \pm 0.41 |
| 5 \pm 0.8 | 0.35 \pm 0.33 | 0.59 \pm 0.25 | 0.64 \pm 0.34 |
| 8 \pm 0.9 | 0.49 \pm 0.39 | 0.48 \pm 0.19 | 0.78 \pm 0.37 |
| 12 \pm 1.1 | 0.43 \pm 0.41 | 0.57 \pm 0.31 | 0.81 \pm 0.35 |

DISCUSSION

The RECIST (RR) method was a quantitative and objective method for monitoring the solid tumors within this cohort of patients, and significantly different from patients with stability or progression of the disease. The RR values of the focus group showed a good decrease, represented by an exponential curve independent of the initial size of the lesions because the changes were calculated relative to the initial tumor size value. That curve was similar to the decreases calculated according to the WHO (RR_{WHO}) method and to our proxy measurement of lesion volume, RR_V .

This finding suggests that any of the three methods allows a similar objective assessment of the response to chemotherapy, which is corroborated by the high

Table 4. Intra-class correlation and number of necessary lesions that do not increase the variance of the metric by more than 10% (n^*) in the different cancer types

| Cancer type | Intra-class correlation | n^* |
|----------------------|-------------------------|-------|
| Renal cell carcinoma | 0.569 | 4.132 |
| Gastric cancer | 0.825 | 1.720 |
| Colon cancer | 0.872 | 1.264 |
| Lung cancer | 0.713 | 2.790 |

correlations among the counts. Despite this, there was a large residual variation in the response, which can be explained by the individual variability in the behavior of the tumors and by the responses to different therapeutic schemes. However, other sources of variability lie in the method's inherent problems that can complicate the assessment. For example, the number of target lesions was not the same in all the patients, which can induce some variability. Other causes of variability may originate from the presence of necrosis that, by favoring fluid accumulation at the center of the tumors, may hinder the evaluation of changes in lesion size, as observed for example in gastrointestinal stromal tumors¹. The difficulty to measure lymph node clusters^{1,7}, as well as the inherent variability attributable to the observer and to the measuring system, could also account for the high residual variability.

The high level of correlation between the metrics was expected because the three parameters are interdependent by definition, although not linearly. However, the Bland-Altman plots showed a curved arc that differed significantly from the curve expected by chance in the comparison between the methods, with an average difference slightly greater than zero. This finding indicates that both the RR and RR_{WHO} methods overestimate the volume changes and that the RR overestimates the RR_{WHO} count. In addition, both the RR and RR_{WHO} counts represent an oversimplification. Therefore, an arc defect, which is similar to that described in correspondence analysis and principal component analysis, arises as a result of the non-linear relationship between all combinations of the metrics¹⁸. This has an important consequence because the diagrams show that the overestimation is greater for the intermediate values. In practice, this could mean that a RR count of 0.5 could indicate a 50% reduction in the lesion size, while an actual RR value of less than 0.5 could still indicate a better response.

We found no difference in the response between the different types of tumors and that all of the methods showed a tumor regression. This can be explained by the fact that the sources of variation of the differential responses may have been homogenized, considering that we were dealing with differently aged patients, with different drugs, and with different susceptibilities to those drugs. In addition, another explanation could be that the counts used consider only the changes in size as an expression of the level of

cell death of the tumor, independently of the cause. This finding is consistent with the model used in the exponential regressions applicable to the study of cell population decay¹⁹.

We also found differences in the tumor behavior between the different types of lesions for both the overall analysis and for the analysis of individual tumors. The graphs show the most pronounced effect on colon, gastric, and renal cell cancers, and they clearly show a less pronounced decrease in the size of the lymph node lesions. This may be a methodological artifact because the RECIST 1.0 method considers the largest diameter of the lymph nodes, which is an incorrect parameter that was fortunately fixed in the later versions of the RECIST method^{5,6}. In tumoral adenopathy, the first attribute is a change in the tumor shape that is characterized by an increase of the short diameter, causing the tumor to lose its natural oval form and acquire a more rounded shape. Moreover, in the case of a response to chemotherapy in patients with lymph node involvement, the lymph node returns to its normal size but does not disappear as the other lesions do (i.e. lung or hepatic nodules). Therefore, the adenopathies have a limited minimum size different from zero. This in itself produces a less pronounced decline of the lesion size. The RECIST 1.1 method requires the monitoring of the short diameter and considers a minimum diameter > 1.5 cm in the case of adenopathies. Unfortunately, this novel criterion corrects the first problem but not the second, so the response curve distortion will likely continue to be present.

The intra-class correlations found in this study were high, between 0.57 and 0.87, suggesting a minimum requirement of between 2-4 in the number of lesions to be followed. This was in the range proposed by Mazumdar, et al.¹⁵, who suggested that the number of lesions to be measured, independent of the total number of lesions, was between 3-6. This result is also consistent with mathematical simulations assessing the necessary number of lesions that adequately describes the response of solid tumors to treatment²⁰. Here, if the behavior of the lymph node lesions is likely a methodological artifact, it also artificially reduces the intra-class correlation and results in an overestimation of the minimum number (n^*) of target lesions to follow for an adequate representation of the response. If this was the case, the n^* 's requirement might actually be even lower. This is consistent

with the results of Zacharia, et al.¹², who showed that only a few lesions (one or two) were sufficient to obtain a good estimate of the response of the liver metastases from colorectal cancer. This result is also consistent with our own estimation of the intra-class correlation in this cancer (0.87), which was the highest and suggested an $n^* = 1.26 \approx 2$ (for safety).

Therefore, our analysis portrays some positive and some negative attributes of the response assessment methods. Among the negative aspects, our study highlights the fact that the currently used metrics are misrepresenting the changes in the lesion volume. Indeed, there is a non-linear overestimation of the counts, more pronounced in the intermediate counts. The inclusion of lymphadenopathy among the target lesions probably introduces a distortion in the evaluation of the response and influences the determination of the minimal number of lesions to follow. The positive aspect highlighted by our study is the fact that quantitative counts allow an objective evaluation of the chemotherapy response of solid tumors. Moreover, our study shows that the response is homogeneous, although it is largely variable, between different types of tumors. The intra-class correlations in the studied cancers were high. These two aspects imply two practical clinical benefits: (i) the monitoring of the patients may be performed with a low number of lesions, and (ii) the evolution of the lesions may be compared with the pattern of good response shown in table 2. For example, a patient with an initial RECIST value of 87 who two months after chemotherapy presents a RECIST value of 50 (and in consequence a RR of 50/87 [0.57]) can be categorized as a good responder. Indeed, this RR value can be compared with the expected value at five months, which is 0.59. Therefore, this patient shows a response that is close to the expected value of a good response and perhaps being 3.5% better than the expected RR value (expected/observed = $0.59/0.57 = 1.035$). Another patient with an initial RECIST of 106, but a RECIST of 87 at three months ($RR = 84/106 = 0.82$) shows a poor response (expected/observed = $0.64/0.82 = 0.78$), considering that the RR value of this patient is 22% worse than the expected RR value. These examples demonstrate how the metrics used in this study could be highly beneficial for monitoring the treatment response of the lesions, especially during the first months of treatment.

DECLARATION OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Suzuki C, Jacobson H, Hatschek T, et al. Radiologic measurements of tumor response to treatment: practical approaches and limitations. *Radiographics*. 2008;28:329-44.
2. World Health Organization. WHO handbook for reporting results of cancer treatment. Available at: <http://whqlibdoc.who.int/publications/9241700483.pdf>.
3. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment on Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst*. 2000;92:205-16.
4. Therasse P, Eisenhauer EA, Verweij J. RECIST revisited: A review of validation studies on tumour assessment. *Eur J Cancer*. 2006;42:1031-9.
5. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228-47.
6. Van Persijn Van Meerten EL, Gelderblom H, Bloem JL. RECIST revised: implications for the radiologist. A review article on the modified RECIST guideline. *Eur Radiol*. 2010;20:1456-67.
7. Canals M, Fernández M, Barros A, Barros P, Salman P, Mahave M. Buscando una curva general de respuesta a la quimioterapia basado en la evaluación radiológica con criterio RECIST. *Rev Chil Radiol*. 2009;15:1-4.
8. James K, Eisenhauer E, Christian M, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst*. 1999;91:523-8.
9. Dacchman AH, MacEneaney PM, Adedipe A, Carlin M, Schumm LP. Tumor size on computed tomography scans: is one measurement enough? *Cancer*. 2001;91:555-60.
10. Warren KE, Patronas N, Aikin AA, Albert PS, Balis FM. Comparison of one-, two-, and three-dimensional measurements of childhood brain tumors. *J Natl Cancer Inst*. 2001;93:1401-5.
11. Mazumdar M, Smith A, Schwartz LH. A statistical simulation study finds discordance between WHO criteria and RECIST guideline. *J Clin Epidemiol*. 2004;57:358-65.
12. Zacharia TT, Saini S, Halpern EF, Sumner JE. CT of colon cancer metastases to the liver using modified RECIST criteria: determining the ideal number of target lesions to measure. *Am J Roentgenol*. 2006;186:1067-70.
13. Hillman SL, Ming-Wen A, O'Connell MJ, et al. Evaluation of the optimal number of lesions needed for tumor evaluation using the Response Evaluation Criteria in Solid Tumors: A North Central Cancer Treatment. *J Clin Oncol*. 2009;27:3205-10.
14. Gonen M, Schwartz L, Ford R. Evaluation of number of target lesions to analyze in time to progression by RECIST. *ASCO Annual Meeting Proceedings (Post-Meeting Edition)*. *J Clin Oncol*. 2007;25:6549.
15. Mazumdar M, Smith A, Debroy PP, Schwartz LH. A theoretical approach to choosing the minimum number of multiple tumors required for assessing treatment response. *J Clin Epidemiol*. 2005;58:150-3.
16. Suzuki C, Blomquist L, Sundin A, et al. The initial change in tumor size predicts response and survival in patients with metastatic colorectal cancer treated with combination chemotherapy. *Ann Oncol*. 2012;23:948-54.
17. Altman DG, Bland JM. Measurement in Medicine: The analysis of method comparison studies. *The Statistician* 1983; 32: 307-317.
18. Rencher AC. *Methods in Multivariate Analysis*, 2nd Ed. New York: Wiley Interscience. J Wiley & Sons, Inc., 2002.
19. Bushong SC. Radiobiología molecular y celular. In Bushong SC, editor. *Manual de Radiología para Técnicos*. Madrid: Esxta Ed. Harcourt Brace, 1998:449-63.
20. Moskowitz C, Jia X, Schwartz L, Gonen M. A simulation study to evaluate the impact of the number of lesions measured on response assessment. *Eur J Cancer*. 2009;45:300-10.