

## Volumetric evaluation of therapy response in patients with lung metastases. Preliminary results with a computer system (CAD) and comparison with unidimensional measurements

### *Valutazione volumetrica della risposta alla terapia in pazienti con metastasi polmonari. Prime esperienze con un sistema di misurazione computerizzato (CAD) e confronto con misurazioni unidimensionali*

F. Fraioli<sup>1</sup> • L. Bertoletti<sup>1</sup> • A. Napoli<sup>1</sup> • F.A. Calabrese<sup>1</sup> • R. Masciangelo<sup>2</sup> • E. Cortesi<sup>3</sup>  
C. Catalano<sup>1</sup> • R. Passariello<sup>1</sup>

<sup>1</sup>Dipartimento di Scienze Radiologiche, Università degli Studi di Roma “La Sapienza”, Viale Regina Elena 324, I-00161 Roma, Italy

<sup>2</sup>Sezione di Statistica, <sup>3</sup>Oncologia Medica, Dipartimento di Medicina Sperimentale e Patologia, Università degli Studi di Roma “La Sapienza”, Roma, Italy

Correspondence to: F. Fraioli, Tel.: +39-06-4455602, Fax: +39-06-490243, e-mail: francesco.fraioli@uniroma1.it

Received: 13 July 2005 / Accepted 18 October 2005 / Published online: 11 April 2006

#### Abstract

**Purpose.** The aim of this study was to assess the efficacy of a Computer-Aided Detection (CAD) system in the identification of lung metastases and to compare the volumetric CAD measurements with unidimensional observer measurements in the evaluation of treatment response in oncology patients.

**Materials and methods.** Two observers (A and B) evaluated nine patients undergoing lung computed tomography (CT) just before and immediately after treatment with chemotherapy. Multislice CT scans were performed before and after the injection of contrast material with a high-resolution protocol (collimation 4x1 mm, 100 mAs, 120 kV). Response Evaluation Criteria in Solid Tumours (RECIST) criteria were used to consider the disease as stable, increased or decreased. Subsequently, target lesions (most significant lesions identified before and after chemotherapy) were evaluated with a computerised system (CAD) to establish volumetric measurements. Observers' unidimensional measurements and CAD volumetric measurements were analysed for comparison.

**Results.** Twenty-four nodules (diameter: 5–18 mm in the first study and 4–20 mm in the follow-up study) were included. Observers agreed in the assessment of therapy response in 21 nodules: eight were considered to have increased in size, and 13 were judged stable. Observer and CAD measurements disagreed in three nodules: two were considered stable by radiologists and increased by CAD; one was considered increased by radiologists and stable by CAD.

As regards patient response, radiologists disagreed in two cases. CAD and observers did not agree in one case.

**Conclusions.** Our preliminary data suggest that volumetric measurements can modify the diagnostic and therapeutic evaluation of oncology patients under chemotherapy.

CAD volumetric measurements allow an easy and objective evaluation, reducing interobserver variability in the evaluation of chemotherapy response.

#### Riassunto

**Obiettivo.** Valutare l'efficacia di un sistema computerizzato (Computer Aided Detection, CAD) nell'identificazione delle metastasi polmonari e confrontare le misurazioni volumetriche effettuate dal sistema, rispetto a quelle unidimensionali operate dagli osservatori, per valutare la risposta alla terapia, in pazienti studiati per follow-up oncologico.

**Materiali e metodi.** Due osservatori (A e B) hanno valutato 9 pazienti con esame TC del torace sia immediatamente prima, che dopo, esser stati sottoposti a chemioterapia. Tutti gli esami sono stati eseguiti prima della somministrazione di MdC, con un protocollo di TC multiestrato (TCMS) ad alta risoluzione del torace (collimazione 4x1 mm, 100 mAs, 120 Kv). Per il giudizio di miglioramento, peggioramento e stabilità di malattia sono stati adottati i criteri RECIST. Successivamente, i noduli, identificati come lesioni target (noduli più significativi presenti sia prima che dopo chemioterapia), sono stati rivalutati mediante un sistema computerizzato in grado di effettuare il calcolo del volume della lesione. Sono stati così comparati, in termini di risposta alla terapia, i risultati ottenuti dagli osservatori con misurazioni su un piano trasversale e quelli del CAD con misurazione volumetrica.

**Risultati.** Ventiquattro noduli (DT: 5–18 mm nel primo studio e 4–20 mm nel follow-up) sono stati inclusi in questa valutazione. Gli osservatori hanno espresso un giudizio di risposta alla terapia sovrapponibile in 21 noduli: 8 considerati aumentati e 13 stabili. Le misurazioni degli osservatori e le misurazioni automatiche fornite dal sistema CAD differivano in 3 noduli: due considerati stabili dai radiologi e aumentati dal CAD e uno considerato aumentato dai radiologi e stabile dal CAD. Considerando la risposta per paziente, una discordanza tra gli osservatori si è verificata in due casi. In un paziente si è evidenziato un diverso giudizio tra gli osservatori e il CAD.

**Conclusioni.** I dati preliminari del nostro studio suggeriscono che le misurazioni volumetriche possono modificare il giudizio diagnostico e quindi terapeutico nella valutazione dei pazienti oncologici in trattamento. Le misurazioni volumetriche effettuate

**Key words** Multislice CT • Thorax • Lung metastases • CAD

*dal CAD consentono con facilità una misurazione oggettiva, riproducibile ed automatica in termini volumetrici, riducendo, quindi, la variabilità interosservatore espressa in questo ed in altri studi nella valutazione della risposta al trattamento dei pazienti oncologici.*

**Parole chiave** TC spirale multistrato • Torace • Metastasi polmonari • CAD

## Introduction

The role of diagnostic imaging in oncology has expanded in the past few years, both in identification and staging of primary disease and in reevaluation of patients undergoing chemotherapy or radiotherapy [1, 2]. Computed tomography (CT) and, in particular, multislice technology, has long been considered a highly sensitive method for identification of lung metastases and evaluation of response to therapy during oncologic follow-up examinations; in fact, its high spatial resolution and high contrast resolution at the level of the lung parenchyma make it an essential diagnostic tool in oncologic patients undergoing chemotherapy.

The international criteria that define oncologic response were initially established in the 1980s by the World Health Organisation (WHO) [3] and recently modified as Response Evaluation Criteria in Solid Tumours (RECIST) criteria [4]. RECIST criteria only take into consideration the largest diameter of the significantly more important lesions (five for each organ; up to ten per patient in the case of multiorgan secondary lesions) and subdivide response to therapy into four categories: stable disease, when there is a <30% reduction or <20% increase in maximum diameter or in the sum of diameters of more than one lesion; partial response, when there is a >30% reduction; progressive disease, when there is a >20% increase; complete response, in the case of complete tumour disappearance [4].

Although they are simple and easy to use, these criteria present some limitations in clinical practice, namely: difficulty in measuring the lesion (extremes of the lesion in the axial plane); different slice thickness used in successive exams (often performed in different facilities or with different devices); ill-defined lesion margins; type of measurement performed (manual or electronic); selection of the target lesions. A consequence of these limitations is often high intra- and interobserver variability, with a highly variable judgement of response to therapy both between the different observers who evaluate follow-up studies and for the same observer reevaluating previous exams [5–8]. In science, this variability is often avoided by consensus: readings by two or more observers, and calculation of the mean among all readings (or, in this case, among measurements); in practice, however, this is often not possible.

One alternative proposed in identification of lung nodules and their measurement is the utilisation of computerised reading devices, which map the lesions and measure their volume. At present, these computer-aided detection (CAD)

## Introduzione

*Il ruolo della diagnostica per immagini in oncologia è andato ampliandosi negli ultimi anni, sia per l'identificazione e stadiazione della neoplasia primitiva, sia nella rivalutazione dei pazienti in corso di trattamenti chemio- o radioterapici [1, 2]. La TC, ed in particolare la nuova tecnologia multistrato, è già da tempo considerata tecnica di elevata sensibilità nella identificazione delle metastasi polmonari e nella valutazione della risposta alla terapia nel corso dei follow-up oncologici; infatti, l'elevata risoluzione spaziale, associata alla elevata risoluzione di contrasto a livello del parenchima polmonare, ne consentono un essenziale strumento diagnostico in pazienti oncologici in corso di chemioterapia.*

*I criteri internazionali che definiscono la risposta oncologica sono stati inizialmente definiti negli anni '80 dalla WHO [3] e recentemente modificati come criteri RECIST [4]. I criteri RECIST prendono in considerazione unicamente il diametro maggiore delle lesioni significativamente più importanti (5 per organo; fino ad un massimo di 10 per paziente, nel caso in cui le lesioni secondarie siano multiorgano), suddividendo la risposta alla terapia in 4 differenti categorie: stabilità di malattia, nel caso in cui vi sia una riduzione inferiore al 30% o un incremento inferiore al 20% del diametro massimo o della somma dei diametri di più lesioni; risposta parziale, nel caso in cui la riduzione sia superiore al 30%; progressione di malattia, nel caso di un incremento superiore al 20%; risposta totale, nel caso di scomparsa totale della lesione [4].*

*Sebbene semplici e di facile utilizzo, i limiti di applicabilità nella pratica clinica sono rappresentati dalla difficoltà di misurazione (estremi massimi di lesione sul piano assiale), dal diverso spessore di strato utilizzato tra esami successivi (spesso eseguiti in strutture o con apparecchiature diverse), dai margini di lesione non ben definibili, dal tipo di misurazione effettuata (manuale o elettronica), dalla scelta delle lesioni "target". Conseguenza di tali limiti è spesso rappresentata da un'alta variabilità intra- ed interosservatore, con un giudizio di risposta alla terapia ampiamente variabile sia tra diversi osservatori che valutano studi successivi, sia nell'osservatore stesso quando si trovi a rivalutare esami precedenti [5–8]. In campo scientifico questa variabilità viene spesso eliminata dal cosiddetto "consensus": letture di due o più osservatori e da una media delle letture (o, in questo caso, misurazioni); in campo pratico ciò spesso non è possibile.*

systems are chiefly applied in the evaluation of lung nodules, both for identification and assessing the growth rate at follow-up [9–11].

The aim of our study was to attempt to integrate these systems in the evaluation of oncologic patients, identify any difference between readings performed by several observers and provide a volumetric evaluation performed with our computerised system.

## Materials and methods

### Patient population and acquisition protocol

Although the study was retrospective, we requested authorisation from the Ethics Committee to collect and process data. During a 14-month period (January 2004 to March 2005), 67 oncologic patients from the Department of Clinical Oncology (age range: 35–62 years) undergoing chemotherapy were evaluated using a four-detector-row multislice CT (MSCT) scanner (Volume Zoom, Forchheim, Germany).

To create a database, we included only patients undergoing chemotherapy with a diagnosis of lung metastasis on the first examination and a CT scan performed during the two chemotherapy cycles. Exclusion criteria were: patients with studies performed with different acquisition protocols; patients who did not have lung metastases at the first study; those who did not have a complete response at the second study (absence of nodular formations within the lung parenchyma). The aim for inclusion was to evaluate both the use of a computerised system and the response to a given chemotherapeutic agent for oncologic patients. Thus, nine patients with breast cancer were selected for our database.

The acquisition protocol we used was the so-called Combi-Mode, with the following technical features: 4x1-mm collimation, 100 mAs, 120 kV. Slice thickness was 1.25 mm, with a 1-mm reconstruction interval and B60 reconstruction algorithm for lung parenchyma analysis; mediastinal evaluation was performed using 5-mm slice thickness reconstructions and B30 algorithm.

The initial lung parenchymal scan was performed only at baseline; then patients were analysed through a complete follow-up exam, with a full chest-abdomen-pelvis and, if necessary, skull CT scan, after administration of a nonionic iodinated contrast agent. For the purposes of this study, patient selection and database creation only took into consideration chest studies performed with the “Combi” protocol. Once acquired, images were transferred to our PACS storage system (Lifeweb, Ferrania Imaging Technologies).

### Selection of metastases

Images of all selected cases were transferred to a dedicated console equipped with a computerised system for identification of lung nodules (LungCheck, R2 Technology, Inc.). Firstly, image reading was done by the two radiologists in consensus comparing on a double screen the first and the follow-up study to identify the target lesions to be measured. A

*Una delle alternative proposte nella identificazione dei noduli polmonari e nella loro misurazione è l'utilizzo di sistemi di lettura computerizzati; essi forniscono, infatti, una mappa delle lesioni ed una loro misurazione volumetrica. Allo stato attuale, tali sistemi, denominati CAD, vengono applicati principalmente alla valutazione dei noduli polmonari, sia nella loro identificazione, sia per la quantificazione del tempo di crescita in studi successivi [9–11].*

*Lo scopo del nostro studio è stato provare a integrare tali sistemi nella valutazione dei pazienti oncologici, identificando eventuali differenze tra una lettura effettuata da diversi osservatori ed una valutazione volumetrica effettuata con il nostro sistema computerizzato.*

## Materiali e metodi

### Popolazione di pazienti e protocollo di acquisizione

*Lo studio da noi proposto è uno studio retrospettivo; per la raccolta dati e per l'elaborazione degli stessi è stata comunque richiesta l'autorizzazione da parte del comitato etico. Durante un periodo di 14 mesi (gennaio 2004, marzo 2005), 67 pazienti oncologici (età compresa tra 35 e 62 anni), provenienti dal Dipartimento di Oncologia Clinica ed in corso di trattamento chemioterapico, sono stati valutati utilizzando un apparecchiatura di TC multiestrato (TCMS) a 4 banchi di rilevatori (Volume Zoom, Forchheim, Germany).*

*Nella creazione del database sono stati inclusi nello studio unicamente i pazienti in corso di chemioterapia, con una diagnosi di metastasi polmonari al primo controllo e con una TC eseguita durante due cicli di chemioterapia. Criteri di esclusione sono stati: pazienti con studi effettuati con protocolli di acquisizione diversi; pazienti che non presentavano metastasi polmonari nel primo studio o con una risposta completa nel secondo studio (assenza di formazioni nodulari nel parenchima polmonare). L'obiettivo per l'inclusione è stato di valutare sia l'utilizzo di un sistema computerizzato, sia la risposta ad un determinato chemio farmaco per i pazienti oncologici.*

*Sono stati così selezionati, nel nostro database, 9 pazienti con tumore della mammella.*

*Il protocollo di acquisizione è stato condotto utilizzando una tecnica definita Combi con le seguenti caratteristiche tecniche: collimazione 4x1 mm, 100 mAs, 120 Kv. Lo spessore di strato utilizzato è stato di 1,25 mm con un intervallo di ricostruzione di 1 mm ed algoritmo di ricostruzione B60 per lo studio del parenchima; la valutazione del mediastino è stata effettuata utilizzando ricostruzioni con spessore di strato di 5 mm e con algoritmo B30.*

*La scansione iniziale sul parenchima polmonare è stata effettuata nelle sole condizioni basali; successivamente le pazienti sono state valutate per un follow-up completo con un esame TC completo di torace-addome-pelvi e, dove richiesto, del cranio, dopo iniezione di mezzo di contrasto organo iodato non ionico. Ai fini dello studio proposto, la selezione dei pazienti e la creazione del database ha preso in considerazione unicamente gli studi eseguiti sul torace con protocollo “Combi”. Le immagini, una volta acquisite, sono*

third radiologist, not included in the case evaluation, created the database, recording in a table the site of target lesions selected by the radiologists. One week later, the two radiologists independently evaluated the size of the lesions selected based on the maximum transverse axis of each lesion. In order to reduce any intra- and interobserver variability, nodule images were enlarged by 150%–200%.

The third radiologist then measured (based on the numerical measurement provided by each observer) the percentage difference between the single lesions, using the RECIST criteria, and providing a general opinion on disease improvement, worsening or stability, considering both response of the single nodules and response for each patient, summing the size of every single nodule. In the case of disagreement in the evaluation of response, a second reading was performed in consensus. The three radiologists in consensus evaluated volumetric measurements performed by the computerised system and response to therapy both for each nodule and for each patient. Thus, 24 metastases were considered in the first reading and 24 in the second.

### Statistical analysis

In the evaluation of response to treatment, the RECIST criteria were taken into consideration (both for single nodules and for each patient). In agreement with volumetric measurements, partial response was defined as >65% reduction; progressive disease was defined as one having >73% volume increase; stable disease was defined as one having <73% increase and <65% reduction (Table 1) [12]. Reading

state trasferite sul nostro sistema di archiviazione PACS (Lifeweb, Ferrania Imaging Technologies).

### Selezione delle metastasi

Tutti i casi selezionati sono stati trasferiti su una console dedicata fornita di un sistema computerizzato di identificazione dei noduli polmonari (LungCheck, R2 Technology, Inc). In un primo momento la lettura dei casi è avvenuta in consensus tra due radiologi, ponendo a confronto su un doppio monitor il primo studio ed il follow-up, per identificare le lesioni target su cui successivamente effettuare le misurazioni. Un terzo radiologo, non incluso nella valutazione dei casi, ha creato il database, registrando su una tabella la sede delle lesioni target scelte dai radiologi.

Successivamente, a distanza di una settimana, i due radiologi, singolarmente, hanno valutato le dimensioni delle lesioni prescelte secondo l'asse massimo trasversale di ogni lesione; per ridurre l'eventuale variabilità di misurazione intra ed interosservatore, le immagini dei noduli sono state ingrandite del 150%–200%.

Il terzo radiologo ha quindi calcolato (sulla risposta numerica di misurazione fornita da ciascun osservatore) la percentuale di differenza tra le singole lesioni, utilizzando i criteri RECIST e fornendo quindi un giudizio complessivo di miglioramento, peggioramento e stabilità di malattia, considerando sia la risposta per singoli noduli, sia la risposta per paziente, effettuando la somma delle dimensioni di ogni singolo nodulo. Nei casi di disaccordo nella valutazione della risposta è stata effettuata una seconda lettura in consensus.

**Table 1** Treatment response categories

Category	Unidimensional criteria <sup>a</sup>	Volumetric criteria <sup>b</sup>
Complete response	Tumour disappearance	Tumour disappearance
Partial response	>30% reduction in diameter	>65% reduction in volume
Stable disease	Size between that for partial and progressive disease	Size between that for partial and progressive disease
Progressive disease	>20% increase in in diameter	>73% increase in volume

<sup>a</sup>Based on Response Evaluation Criteria in Solid Tumours (RECIST) guidelines. In patients with multiple lesions, individual maximum diameters were summed

<sup>b</sup>Based on RECIST guidelines

**Tabella 1** Categorie di risposta al trattamento

Categorie	Criteri unidimensionali <sup>a</sup>	Criteri volumetrici <sup>b</sup>
Risposta completa	Scomparsa delle lesioni	Scomparsa delle lesioni
Risposta parziale	Riduzione >30% del diametro massimo	Riduzione >65% del volume
Stabilità di malattia	Dimensioni comprese tra risposta parziale e progressione di malattia	Dimensioni comprese tra risposta parziale e progressione di malattia
Progressione di malattia	Aumento >20% del diametro massimo	Aumento >73% del volume

<sup>a</sup>Secondo i criteri RECIST. Nei pazienti con lesioni multiple, il diametro massimo di ogni singola lesione viene sommato

<sup>b</sup>Criteri RECIST

results were then entered into a spreadsheet, and interobserver agreement and measurement conformity by a single observer with the RECIST criteria and CAD volumetric measurements were evaluated using statistical analysis software (SPSS 8.0 for Windows; SPSS; Chicago, IL, USA). Thus, kappa-type statistics were used to determine very poor ( $k < 0.01$ ), poor ( $k 0.01-0.20$ ), moderate ( $k 0.21-0.40$ ), good ( $k 0.41-0.60$ ), very good ( $k 0.61-0.80$ ) and excellent ( $k 0.81-1.00$ ) agreement [13]. For each test performed, a limit of  $p < 0.005$  was defined as statistically significant.

## Results

Twenty-four nodules were included in the evaluation. The diameter of nodules studied ranged between 5 mm and 18 mm in the first study and between 4 mm and 20 mm at follow-up. Table 2 shows the total of results obtained in the different measurements by observers A and B and by CAD. Although measurements were different for most examined nodules, as can be seen in the table, considering the RECIST criteria, which allow for moderate variability, judgement of response to therapy coincided in most cases, both in evaluation of single nodules (Table 2) and for each patient (Table 3) (Fig. 1).

Evaluation of response to therapy for each nodule provided different results in three cases: a nodule considered reduced by observer A was defined as stable by observer B (Fig. 2); two nodules considered as increased by observer A were defined as stable by observer B; considering the response of each patient, there was interobserver disagreement in two cases – one was considered as increased by one observer and stable by the other and one case considered as disease reduction by one observer was evaluated as stable by the other. To compare response to therapy measured by the observers with that measured by the CAD system, a consensus reading was performed by the two observers, who reevaluated three nodules.

Compared with the consensus reading, the automatic measurements performed by the CAD system were different for three nodules: two were considered stable by the radiologists and increased by CAD; one was considered increased by the radiologists and stable by CAD (Fig. 3). Considering results for each patient, there was one case of disagreement between the radiologists, who stated the patient was stable, and CAD, which revealed disease progression. Consensus results between the observers and between CAD and the observers are shown in Table 4.

## Discussion

An objective evaluation of response to therapy in oncologic patients is an important goal in routine radiological activity both concerning clinical decisions on patient treatment and trials on new chemotherapeutic agents [2]. In the past few years, assessment and follow-up of oncologic patients has

*Successivamente, i tre radiologi in consensus hanno valutato le dimensioni volumetriche effettuate dal sistema computerizzato ed hanno valutato la risposta alla terapia, anche in questo caso sia per nodulo che per paziente. Sono state così prese in considerazione 24 metastasi nella prima lettura e 24 nella seconda lettura.*

## Analisi statistica

*Nella valutazione della risposta al trattamento sono stati presi in considerazione i criteri RECIST (sia per i noduli singoli, sia per paziente). In accordo con le misurazioni volumetriche, la risposta parziale è stata definita come una riduzione superiore al 65%; la progressione di malattia come aumento del volume superiore al 73%; la stabilità di malattia come aumento inferiore al 73% e diminuzione inferiore al 65% (Tabella 1) [12].*

*I risultati delle letture sono stati inseriti in un foglio di lavoro excel e valutati, utilizzando un programma statistico (SPSS 8.0 per Windows; SPSS; Chicago, III), in termini di accordo inter-osservatore e di concordanza tra le misurazioni del singolo osservatori con i criteri RECIST e le misurazioni volumetriche del CAD.*

*È stata così utilizzata una statistica K per determinare scarsa ( $k < 0,01$ ), bassa ( $k = 0,01-0,20$ ), discreta ( $k = 0,21-0,40$ ), buona ( $k = 0,41-0,60$ ), sostanziale concordanza ( $k = 0,61-0,80$ ) e concordanza quasi perfetta ( $k = 0,81-1,00$ ) [13]. Per ogni test statistico effettuato è stato definito come parametro di significatività un limite di  $p < 0,005$ .*

## Risultati

*Ventiquattro noduli sono stati inclusi in questa valutazione. Il diametro dei noduli valutati è risultato compreso tra 5 e 18 mm nel primo studio e 4 e 20 mm nel follow-up. La Tabella 2 evidenzia il totale dei risultati ottenuti dalle diverse misurazioni per il lettore A, B e per il CAD. Sebbene, come si evince dalla Tabella 2, le misurazioni risultino diverse nella maggior parte dei noduli esaminati, prendendo in considerazione i criteri RECIST, che consentono un discreto margine di variabilità, il giudizio di risposta alla terapia è risultato in molti casi sovrapponibile tra gli osservatori sia per la valutazione dei singoli noduli (Tabella 2), che per paziente (Tabella 3) (Fig. 1).*

*La valutazione della risposta alla terapia per nodulo evidenzia differenti risultati in tre casi: un nodulo, considerato ridotto per l'osservatore 1, è stato definito stabile per l'osservatore 2 (Fig. 2); due noduli, considerati aumentati per osservatore 1, sono stati definiti stabili per l'osservatore 2; considerando la risposta per paziente, una discordanza tra gli osservatori è stata evidenziata in due casi: uno considerato aumentato per un osservatore e stabile per l'altro, mentre un caso considerato come riduzione di malattia per un osservatore, è stato valutato come stabile per l'altro osservatore.*

*Per correlare la risposta alla terapia rilevata tra gli osservatori e quella del sistema CAD è stato effettuato un consensus*

**Table 2** Results of dimensional evaluation for single nodule

A1	B1	V1	A2	B2	V2	A	B	Consensus	V
7	6	98	12	12	632	3	3	3	3
7	7	92	11	13	895	3	3	3	3
6	5	128	6	4	68	2	2	2	2
6	6	108	6	6	136	2	2	2	2
10	12	434	13	13	1038	3	2	3	3
7	7	104	9	9	354	3	3	3	3
7	8	164	11	10	523	3	3	3	3
16	18	1022	18	20	1784	2	2	2	3
7	7	172	7	7	155	2	2	2	2
6	6	98	6	6	80	2	2	2	2
8	8	208	8	8	211	2	2	2	2
8	9	356	9	9	298	2	2	2	2
10	10	184	5	9	94	1	2	2	2
15	14	349	18	18	436	3	3	3	2
8	6	147	8	6	194	2	2	2	2
13	13	422	12	12	334	2	2	2	2
6	5	93	10	10	582	3	3	3	3
7	6	109	13	13	418	3	3	3	3
7	7	57	6	6	59	2	2	2	2
11	11	319	11	10	320	2	2	2	2
12	12	416	10	10	299	2	2	2	2
8	8	246	11	9	433	3	2	3	3
7	7	144	9	9	384	3	3	3	3
8	8	316	8	9	649	2	2	2	3

A1-B1-A2-B2, nodule maximum diameters (mm) calculated by the first (A) and second (B) radiologist regarding the first (1) and the follow-up (2) computed tomography (CT) scan; V1-2, nodule volume (cc) calculated by the Computer-Aided Detection (CAD) system regarding the first (1) and the follow-up (2) CT scan; A-B-Consensus-V, therapy response: 1=partial response; 2=stable disease; 3=progressive disease

**Tabella 2** Risultati della valutazione dimensionale per singolo nodulo

A1	B1	V1	A2	B2	V2	A	B	Consensus	V
7	6	98	12	12	632	3	3	3	3
7	7	92	11	13	895	3	3	3	3
6	5	128	6	4	68	2	2	2	2
6	6	108	6	6	136	2	2	2	2
10	12	434	13	13	1038	3	2	3	3
7	7	104	9	9	354	3	3	3	3
7	8	164	11	10	523	3	3	3	3
16	18	1022	18	20	1784	2	2	2	3
7	7	172	7	7	155	2	2	2	2
6	6	98	6	6	80	2	2	2	2
8	8	208	8	8	211	2	2	2	2
8	9	356	9	9	298	2	2	2	2
10	10	184	5	9	94	1	2	2	2
15	14	349	18	18	436	3	3	3	2
8	6	147	8	6	194	2	2	2	2
13	13	422	12	12	334	2	2	2	2
6	5	93	10	10	582	3	3	3	3
7	6	109	13	13	418	3	3	3	3
7	7	57	6	6	59	2	2	2	2
11	11	319	11	10	320	2	2	2	2
12	12	416	10	10	299	2	2	2	2
8	8	246	11	9	433	3	2	3	3
7	7	144	9	9	384	3	3	3	3
8	8	316	8	9	649	2	2	2	3

A1-B1-A2-B2, diametro massimo dei noduli espresso in mm dal primo (A) e dal secondo (B) radiologo, relativo al primo esame TC (1) e al successivo follow-up (2); V1-2, volume dei noduli espresso in cc dal sistema CAD relativi al primo esame TC (1) e al successivo follow-up (2); A-B-Consensus-V, risposta alla terapia: 1=risposta parziale, 2=stabilità, 3=progressione

**Table 3** Results of dimensional evaluation for each patient

A1	B1	V1	A2	B2	V2	A	B	Consensus	V
26	24	426	35	35	1731	3	3	3	3
40	45	1724	51	52	3699	3	2	3	3
29	30	834	30	30	744	2	2	2	2
35	24	533	23	27	529	1	2	2	2
21	19	569	20	18	528	2	2	2	2
13	11	202	23	23	1000	3	3	3	3
18	18	376	17	16	379	2	2	2	2
12	12	416	10	10	299	2	2	2	2
23	23	706	18	27	1466	2	2	2	3

*A1-B1-A2-B2*, sum of maximum diameter of nodules (mm) for each patient calculated by the first (A) and second radiologist (B) regarding the first (1) and the follow up (2) computed tomography (CT) scan; *V1-2*, sum of nodule volumes (cc) for each patient calculated by the Computer-Aided Detection (CAD) system regarding the first (1) and the follow-up CT scan (2); *A-B-Consensus-V*, therapy response: 1=partial response; 2= stable disease; 3=progressive disease

**Tabella 3** Risultati della valutazione dimensionale per ciascun paziente

A1	B1	V1	A2	B2	V2	A	B	Consensus	V
26	24	426	35	35	1731	3	3	3	3
40	45	1724	51	52	3699	3	2	3	3
29	30	834	30	30	744	2	2	2	2
35	24	533	23	27	529	1	2	2	2
21	19	569	20	18	528	2	2	2	2
13	11	202	23	23	1000	3	3	3	3
18	18	376	17	16	379	2	2	2	2
12	12	416	10	10	299	2	2	2	2
23	23	706	18	27	1466	2	2	2	3

*A1, B1, A2, B2*, somma dei diametri massimi dei noduli (per ogni paziente) espresso in mm calcolati sulla base della lettura del primo (A) e dal secondo (B) radiologo, relativa al primo esame TC (1) e al successivo follow-up (2); *V1-2*, somma dei volumi dei noduli (per ogni paziente) espresso in cc dal sistema CAD relativi al primo esame TC (1) e al successivo follow-up (2); *A-B-Consensus-V*, risposta alla terapia: 1=risposta parziale, 2=stabilità, 3=progressione

**Table 4** Interobserver agreement (observers 1 and 2) and agreement between the two radiologists in consensus and volumetric measurements provided by the Computer-Aided Detection (CAD) system (Consensus vs. CAD)

	Observer 1 vs. 2	Consensus vs. CAD
Nodules	0.75	0.75
Patients	0.59	0.69

**Tabella 4** Accordo interosservatore (osservatori 1 e 2) e accordo tra i due radiologi in consensus e le misurazioni volumetriche fornite dal CAD (consensus vs. CAD)

	Oss. 1 vs. 2	Consensus vs. CAD
Noduli	0,75	0,75
Pazienti	0,59	0,69

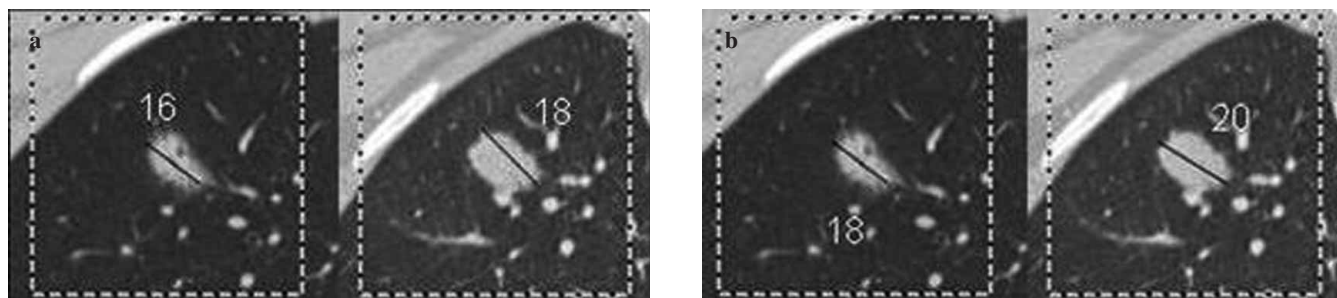
become an integral part of most exams performed by radiologists, at least in hospital. Among the different reasons for this increase in the number of exams, two are fundamental: first, the availability of more aggressive therapies with anti-cancer drugs that are increasingly more effective against secondary formations, resulting in better patient outcome and

tra gli osservatori che hanno così rivalutato 3 noduli. Rispetto all'accordo tra gli osservatori le misurazioni automatiche fornite dal sistema CAD differivano in 3 noduli: due considerati stabili dai radiologi e aumentati dal CAD e uno considerato aumentato dai radiologi e stabile dal CAD (Fig. 3). Considerando i risultati per paziente, in un caso si è avuta una discordanza di valutazione tra i radiologi, per i quali il paziente era stabile, e il CAD, per il quale c'era stata una progressione di malattia. I risultati dell'accordo tra gli osservatori e tra il CAD e gli osservatori sono espressi nella Tabella 4.

## Discussione

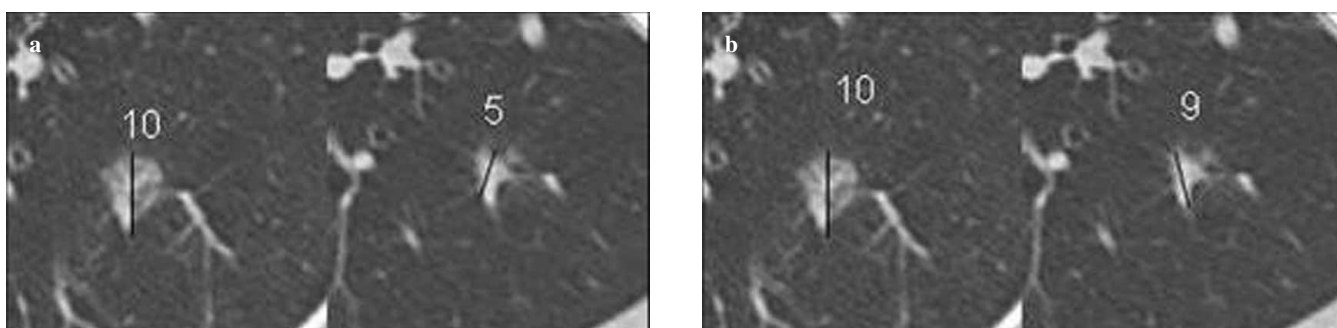
Una valutazione oggettiva della risposta alla terapia nei pazienti oncologici rappresenta un importante obiettivo per il quotidiano lavoro radiologico sia nella decisione clinica sul trattamento di un paziente, sia nelle sperimentazioni di nuove chemioterapie [2].

Negli ultimi anni la valutazione ed il follow-up dei pazienti oncologici è divenuta parte integrante della maggior parte degli esami effettuati dal radiologo, almeno in ambito ospedaliero.



**Fig. 1a,b** Metastatic lung nodule in right upper lobe, evaluated with two multislice computed tomography (MSCT) studies before and after chemotherapy. Agreement between the first (a) and second observer (b) in evaluation of therapy response. Based on Response Evaluation Criteria in Solid Tumours (RECIST) criteria, both radiologists considered the disease as stable.

*Fig 1a,b* Nodulo polmonare di natura secondaria localizzato a livello del lobo superiore del polmone di destra, valutato con due esami TCMS prima e dopo chemioterapia. Concordanza tra il primo (a) e il secondo (b) osservatore nella valutazione della risposta alla terapia. In accordo con i criteri RECIST, entrambi i radiologi hanno classificato la patologia come stabile.



**Fig. 2a,b** Lung metastases in left lower lobe. A case of disagreement between the two observers: the first (a) considered the nodule decreased in size while the second (b) considered the nodule as stable after therapy. At consensus reading, it was judged as stable.

*Fig 2a,b* Metastasi polmonare localizzata a livello del lobo inferiore del polmone di sinistra. Un caso di discordanza: il primo osservatore (a) ha giudicato la lesione come diminuita; il secondo (b) come stabile. Il consensus, effettuato in un secondo momento, ha classificato il nodulo come stabile dopo terapia.

marked improvement in survival and quality of life; second, greater availability of increasingly sophisticated CT scanners, which results in a rise in requests for radiological examinations. These fundamentals result in a need to evaluate an increasing number of CT exams and in particular to study, even at short time intervals, exams of oncologic patients undergoing chemotherapy or actinic therapy.

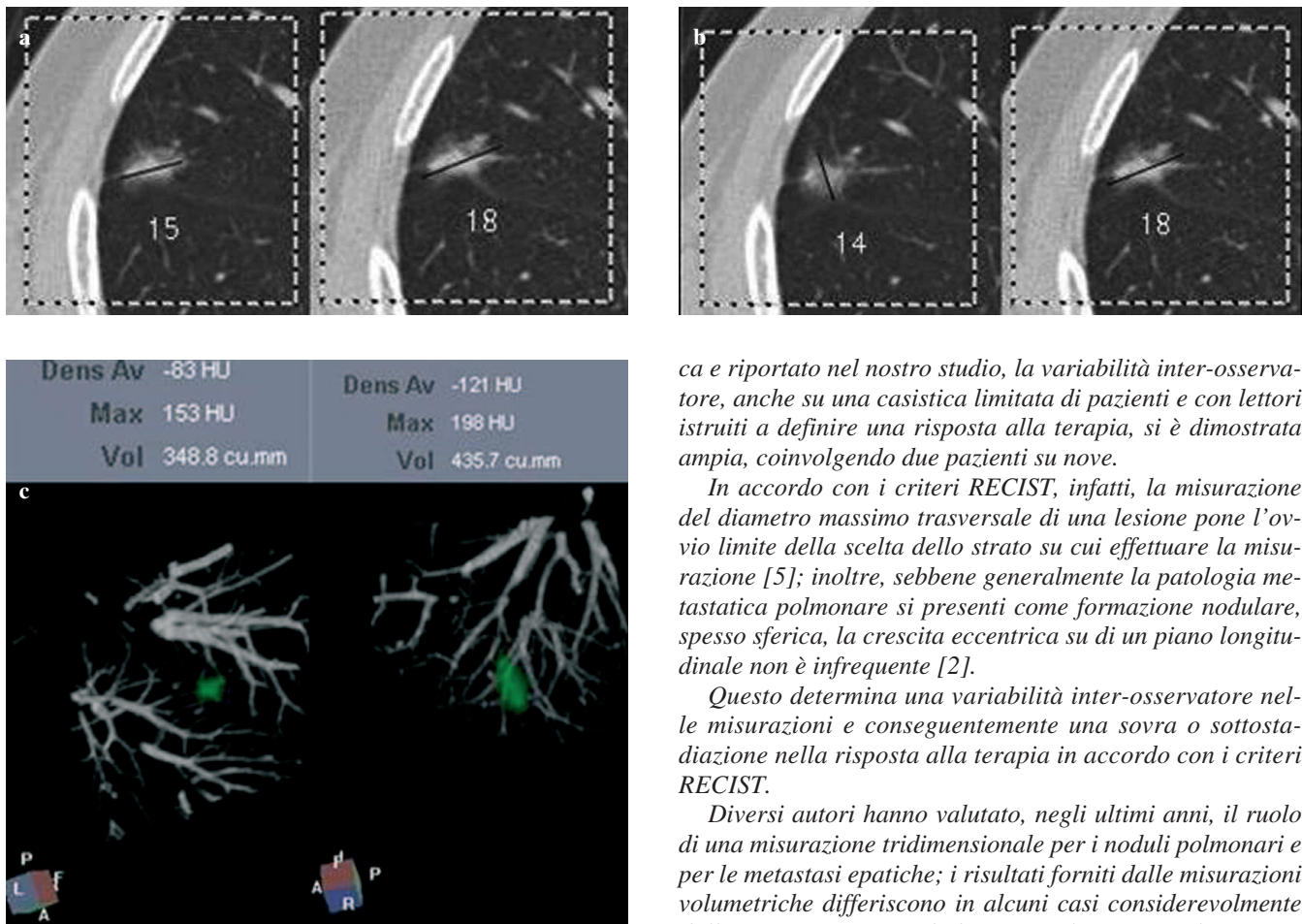
In spite of the growing diagnostic reliability of latest-generation scanners in diagnosis and staging, the large number of exams to be read and compared with previous exams – often performed with different imaging techniques – raises the issue of objective variability of opinions [7]. One system adopted by our department for evaluating all oncologic patients and for reducing intra- and interobserver variability is to use standard protocols, reproducible and comparable in successive studies. Nevertheless, as revealed in clinical practice and reported in our study, interobserver variability, even on a limited number of patients and with readers asked to define a response to therapy, proved to be wide and, in our study, involved two out of nine patients. In fact, according to RECIST criteria, measurement of the greatest transverse diameter of a lesion involves the obvious limitation of selecting the slice on which to perform the measurement [5]; furthermore, although metastatic lung diseases generally consist of often spherical nodular masses, eccentric growth in the

*Tra le molteplici ragioni di questo incremento nel numero di esami vi sono certamente due fondamentali. Primo, terapie più aggressive, con farmaci antitumorali di maggior successo nelle forme secondarie, con una prognosi per il paziente maggiormente favorevole ed un netto miglioramento della durata e della qualità di vita. Secondo, la maggiore disponibilità di apparecchiature sempre più sofisticate di TC, con un conseguente incremento della richiesta di esami radiologici. Il risultato che ne deriva è rappresentato dalla necessità di valutare un numero crescente di esami TC ed in particolare di valutare, spesso a distanze ravvicinate, esami di pazienti oncologici in corso di terapie chemioterapiche o attiniche. Nonostante la sempre maggior affidabilità diagnostica delle apparecchiature di ultima generazione nella diagnosi e stadiazione, l'elevato numero di esami da dover giornalmente valutare, ponendo a confronto esami precedenti, peraltro spesso eseguiti con tecniche di studio diverse, pone il radiologo di fronte ad una oggettiva variabilità di giudizio [7].*

*Uno dei sistemi adottati nel nostro Dipartimento per la valutazione di tutti i pazienti oncologici e per ridurre la variabilità intra ed interosservatore è di utilizzare protocolli standard, e quindi riproducibili e confrontabili negli studi successivi.*

*Nonostante ciò, così come evidenziato nella pratica clini-*





**Fig. 3a-c** Lung metastases in right middle lobe. Both observers agreed in classifying the nodule as progressive disease according to Response Evaluation Criteria in Solid Tumours (RECIST) criteria (a,b); Computer-Aided Detection (CAD) volumetric measurements classified the disease as stable (c).

**Fig 3a-c** Metastasi polmonare localizzata a livello del lobo medio del polmone di destra. I due radiologi (a,b) concordavano, in accordo con i criteri RECIST, sulla progressione di malattia, mentre le misurazioni volumetriche effettuate dal CAD consideravano stabile la patologia secondaria dopo terapia (c).

longitudinal plane is not uncommon [2]. This leads to inter-observer variability among measurements and, therefore, an over- or underestimation of response to therapy according to RECIST criteria.

Several authors have evaluated the role of three-dimensional (3D) measurements of lung nodules and liver metastases; results provided by volumetric measurements present, in some cases, considerable differences from uni- and bidimensional measurements, in particular when the lesion margins are ill defined [12, 14]. Nonetheless, one limitation of using a method for volumetric measurement is the long time needed to calculate the volume, mainly due to the manual drawing of the lesion edges; furthermore, small or ill-defined lesions are often associated with manual errors in volume calculation [15].

The aim of our study was to apply a computerised system capable of automatically performing a volumetric measurement of a lesion to the evaluation of oncologic patients with

ca e riportato nel nostro studio, la variabilità inter-osservatore, anche su una casistica limitata di pazienti e con lettori istruiti a definire una risposta alla terapia, si è dimostrata ampia, coinvolgendo due pazienti su nove.

In accordo con i criteri RECIST, infatti, la misurazione del diametro massimo trasversale di una lesione pone l'ovvio limite della scelta dello strato su cui effettuare la misurazione [5]; inoltre, sebbene generalmente la patologia metastatica polmonare si presenti come formazione nodulare, spesso sferica, la crescita eccentrica su di un piano longitudinale non è infrequente [2].

Questo determina una variabilità inter-osservatore nelle misurazioni e conseguentemente una sovra o sottostima nella risposta alla terapia in accordo con i criteri RECIST.

Diversi autori hanno valutato, negli ultimi anni, il ruolo di una misurazione tridimensionale per i noduli polmonari e per le metastasi epatiche; i risultati forniti dalle misurazioni volumetriche differiscono in alcuni casi considerevolmente dalle misurazioni uni- e bidimensionali, in particolare quando i margini della lesione si presentano irregolari [12, 14]. Uno dei limiti evidenziati nell'utilizzo di un metodo di misurazione volumetrica è tuttavia l'ampio tempo necessario al calcolo del volume, legato, in particolare, alla traccia manuale dei margini di lesione; inoltre, spesso, lesioni di piccole dimensioni o a margini irregolari si associano ad errori di manualità nel calcolo del volume [15].

Lo scopo del nostro studio è stato di applicare un sistema computerizzato, in grado di effettuare automaticamente il calcolo volumetrico di una lesione, alla valutazione dei pazienti oncologici, con metastasi polmonari già accertate, in corso di chemioterapia e di confrontare i risultati di misurazione ottenuti dal nostro sistema CAD con quelli dei due osservatori.

L'introduzione di software automatici per l'identificazione dei noduli polmonari è una realtà che si sta sempre più affermando nella pratica clinica del radiologo. Questi sistemi hanno dimostrato una ampia affidabilità diagnostica, con valori di sensibilità superiori al 90% e superiori alla valutazione dello stesso radiologo [9-11].

Una delle caratteristiche tecniche di questi sistemi è la valutazione automatica delle dimensioni del nodulo in termini di volume, consentendo quindi di comparare studi diversi; l'utilizzo di tali sistemi computerizzati allo stato attuale è stato limitato alla identificazione dei noduli polmonari e alla valutazione del tempo di crescita. Tale ruolo si è dimostrato estremamente significativo, consentendo di poter mi-

proven lung metastases and undergoing chemotherapy, as well as to compare results of measurements obtained with the CAD system with those provided by the two observers.

Automated software for identification of lung nodules is increasingly being introduced into radiological clinical practice. These systems have good diagnostic accuracy with higher than 90% sensitivity, even better than the radiologist's evaluation [9–11]. One technical features of these systems is automatic assessment of nodule size in terms of volume, which allows for comparison between different studies; at present, use of these techniques is limited to identification of lung nodules and assessment of their growth pattern. This role has proved to be extremely significant, providing better evaluation of volumetrically increased nodules that remain stable in the transverse plane [16].

In our study, we took into consideration interobserver consensus, thus obtaining fair agreement between the response to therapy assessed by the observers and that assessed through the measurement; nevertheless, the observers recorded three cases in which images had to be reevaluated to reach a consensus. This means that interobserver variability plays an important role in establishing therapy success. On the other hand, other authors have shown that volume measurement by a computerised system is reliable and reproducible among different studies [17]. Thus, the possibility of evaluating response to therapy using a computerised technique able to provide volumetric results represents valuable support for the radiologist, as it reduces reading times and provides more objective evaluation, even for lung metastases.

## Conclusions

CAD volumetric measurements easily allow for objective, reproducible and automated volume measurement, thus limiting interobserver variability in evaluation of response to treatment of oncologic patients. Our study presents some limitations, the first being the limited number of patients and the metastases analysed. Therefore, it may be argued that a greater number of patients could invalidate the results obtained in the present study. Furthermore, a different selection of more spherical or irregular metastases could modify results of our volumetric measurement, leading to greater agreement or disagreement between observers and CAD. Thus, our study only presents preliminary results and aimed at assessing the possibility of using computerised techniques even in the evaluation of response to treatment of oncologic patients. A second limitation found in the daily application of these systems is the need to use thin-collimation protocols, indispensable for CAD image processing, which lead to difficult comparisons with exams performed in other centres and with different protocols. For this reason, our study had to be retrospective, performed only on patients who had undergone high-resolution exams that could be analysed by CAD. Two considerations need to be made about this limitation: first is that the need to use thin-slice protocols for correct evaluation of lung nodules has recently been demonstrated

*giorare la valutazione di noduli cresciuti volumetricamente, ma stabili su un piano di misurazione trasversale [16].*

*Nel nostro studio, prendendo in considerazione il consenso tra gli osservatori, si è ottenuta una buona concordanza tra la risposta alla terapia dei lettori e quella fornita da una misurazione volumetrica; nonostante ciò, tra gli osservatori vi sono stati 3 casi in cui le immagini sono state rivalutate per raggiungere un accordo.*

*Ne consegue che la variabilità interosservatore gioca un ruolo importante nello stabilire il corretto andamento di una terapia.*

*Di contro, la valutazione del volume di un sistema computerizzato è stata dimostrata da altri autori essere invece una misura affidabile e quindi riproducibile nella comparazione tra i diversi studi [17].*

*La possibilità di valutare la risposta alla terapia utilizzando un sistema computerizzato, in grado di fornire risultati volumetrici, rappresenta, quindi, un valido ausilio per il radiologo, riducendo il tempo di lettura e fornendo una valutazione maggiormente oggettiva anche per le metastasi polmonari.*

## Conclusioni

*Le misurazioni volumetriche effettuate dal CAD consentono con facilità una misurazione oggettiva, riproducibile ed automatica in termini volumetrici, riducendo, quindi, la variabilità interosservatore nella valutazione della risposta al trattamento dei pazienti oncologici.*

*Il nostro studio presenta alcune limitazioni, prima fra tutte il limitato numero di pazienti e di metastasi valutate. È quindi ipotizzabile che un maggior numero di pazienti possa modificare i risultati ottenuti nel presente lavoro; inoltre, una differente selezione di metastasi, maggiormente sferiche o irregolari, potrebbe variare i risultati di una misurazione volumetrica con una maggior concordanza o discordanza tra gli osservatori ed il CAD.*

*Il presente lavoro offre, quindi, risultati preliminari e si pone, come obiettivo, la possibilità di utilizzo di sistemi computerizzati anche nella valutazione della risposta al trattamento dei pazienti oncologici.*

*Un secondo limite nell'applicazione routinaria di tali sistemi è rappresentato dalla necessità di utilizzare protocolli con collimazioni sottili, indispensabili per l'elaborazione delle immagini da parte del CAD, con difficoltà nell'operare confronti con esami effettuati in altri centri e con protocolli diversi.*

*Ciò ha ristretto il nostro lavoro ad uno studio retrospettivo, effettuato unicamente in pazienti valutati con tecnica ad alta risoluzione e, quindi, analizzabili dal CAD.*

*Riguardo questo problema ci sono due considerazioni: la prima è la necessità di utilizzare protocolli con spessori di strato sottili, per una corretta valutazione dei noduli polmonari, come recentemente dimostrato da un articolo del dr. Zhao et al. [5]; la seconda, riguarda la crescente introduzione, nei maggiori centri diagnostici, di apparecchiature TC multistrato con collimazioni fisse o variabili inferiori al*

by Zhao et al. [5]; the second consideration concerns the increasing use of multislice CT scanners, with collimations that are either fixed or variable to less than 1 mm, that allow for storage on increasingly sophisticated PACS systems of sub-millimetre raw data that can thus be used by CAD systems.

The final limitation is that CAD could be unable to identify all nodules chosen as target lesions. In our study, this did not happen, as CAD identified all nodules selected as target lesions; however, it is still possible to manually add a lesion not identified by CAD on which the system can perform the same volumetric measurement.

Integration of new systems capable of automatically comparing two studies on different screens and providing the percentage of variation of a lesion paves the way for routine use of CAD systems, even in evaluation of lung metastases, as well as representing a valuable second reading in screening programmes and identification of lung nodules.

*mm, con possibilità di archiviare su sistemi PACS sempre più performanti i dati grezzi submillimetrici, utilizzabili quindi anche dai sistemi CAD.*

*Infine, un'ultima limitazione è che il CAD potrebbe non identificare tutti i noduli scelti come lesioni target. Nel nostro studio ciò non si è verificato, avendo il CAD identificato tutte le formazioni nodulari prescelte come lesioni target; nonostante ciò, è possibile aggiungere manualmente una lesione non identificata dal CAD, su cui viene eseguita dal sistema la stessa misurazione volumetrica.*

*L'integrazione di nuovi sistemi in grado di confrontare automaticamente due studi su due diversi monitor e di fornire la percentuale di variazione di una lesione apre quindi il campo all'utilizzo routinario dei sistemi CAD anche nella valutazione delle metastasi polmonari, oltre che come valido secondo lettore nei programmi di screening e di identificazione dei noduli polmonari.*

## References/Bibliografia

- Hopper KD, Singapuri K, Finkel A (2000) Body CT and oncologic imaging. *Radiology*. 215:27–40
- Saini S (2001) Radiologic measurement of tumor size in clinical trials: past, present, and future. *AJR Am J Roentgenol* 176:333–334
- World Health Organization (1979) WHO handbook for reporting results of cancer treatment. WHO offset publication no. World Health Organization, Geneva, Switzerland
- Therasse P, Arbuk SG, Eisenhauer EA et al (2000) New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 92:205–216
- Zhao B, Schwartz LH, Moskowitz CS et al (2005) Pulmonary metastases: effect of CT section thickness on measurement. Initial experience. *Radiology* 234:934–939
- Schwartz LH, Ginsberg MS, DeCorato D et al (2000) Evaluation of tumor measurements in oncology: use of film-based and electronic techniques. *J Clin Oncol* 18:2179–2184
- Thiess P, Ollivier L, Di Stefano-Louineau D et al (1997) Response rate accuracy in oncology trials: reasons for interobserver variability. *J Clin Oncol* 15:3507–3514
- Hopper KD, Kasales CJ, Van Slyke MA et al (1996) Analysis of interobserver and intraobserver variability in CT tumor measurements. *AJR Am J Roentgenol* 167:851–854
- Fraioli F, Catalano C, Almberger M et al (2005) Evaluation of effectiveness of a computer system (CAD) in the identification of lung nodules with low-dose MSCT: scanning technique and preliminary results. *Radiol Med* 109:40–48
- Rubin GD, Lyo JK, Paik DS et al (2005) Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. *Radiology* 234:274–283
- Awai K, Murao K, Ozawa A et al (2004) Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. *Radiology* 230:347–352
- Prasad SR, Jhaveri KS, Saini S et al (2002) CT tumor measurement for therapeutic response assessment: comparison of unidimensional, bidimensional, and volumetric techniques initial observations. *Radiology* 225:416–419
- Landis JR, Koch GG (1977) An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 33:363–374
- Van Hoe L, Van Cutsem E, Vergote I et al (1997) Size quantification of liver metastases in patients undergoing cancer treatment: reproducibility of one-, two-, and three dimensional measurements determined with spiral CT. *Radiology* 202:671–675
- Sohaib SA, Turner B, Hanson JA et al (2000) CT assessment of tumour response to treatment: comparison of linear, cross-sectional and volumetric measures of tumour size. *Br J Radiol* 73:1178–1184
- Revel MP, Bissery A, Bienvenu M et al (2004) Are two-dimensional CT measurements of small noncalcified pulmonary nodules reliable? *Radiology* 231:453–458
- Wormanns D, Kohl G, Klotz E et al (2004) Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. *Eur Radiol* 14:86–92