creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

의학박사 학위논문

Illustration of Impact on Measurement Variability on Response Categorization and Objective Response Rate by Hypothetical Re-assessments

가상적인 재평가를 통해 치료반응 분류 및 객관적 반응률에 대한 측정 변이의 영향 실증

2019년 2월

서울대학교 대학원
의학과 영상의학전공
윤 순 호

A thesis of the Degree of Doctor of Philosophy in Medicine

가상적인 재평가를 통해 치료반응 분류 및 객관적 반응률에 대한 측정 변이의 영향 실증

Illustration of Impact on Measurement Variability on Response Categorization and Objective Response Rate by Hypothetical Re-assessments

Feb 2019

The Department of Radiology,
Seoul National University College of Medicine

Soon Ho Yoon

Illustration of Impact on Measurement Variability on Response Categorization and Objective Response Rate by Hypothetical Re-assessments

가상적인 재평가를 통해 치료반응 분류 및 객관적 반응률에 대한 측정 변이의 영향 실증

지도 교수 구 진 모

이 논문을 의학박사 학위논문으로 제출함
2019년 2월

서울대학교 대학원
의학과 영상의학 전공
윤 순 호

윤순호의 의학박사 학위논문을 인준함
2019년 2월

위 원 장 한 준 구 (인)
부위원장 구 진 모 (인)
위 원 한 서 경 (인)
위 원 김 동 완 (인)
위 원 김 태 정 (인)

Abstract

Purpose: To investigate the measurement variability between assessments of tumor burdens, and its impact on response categorization and the resulting efficacy outcomes which are applicable to oncology trials.

Methods: We built up a hierarchical model of measurement variability using a trial dataset of CT scans. Simulations were then performed using the model in various scenarios 1) to establish the behavior of differences between the first and the second assessments of percent change, 2) to elaborate on the probabilistic nature of decisions about categorization, and 3) to estimate variation in the assessed objective response rate (ORR) by generating an 95% central range of ORR results if the reassessment was performed.

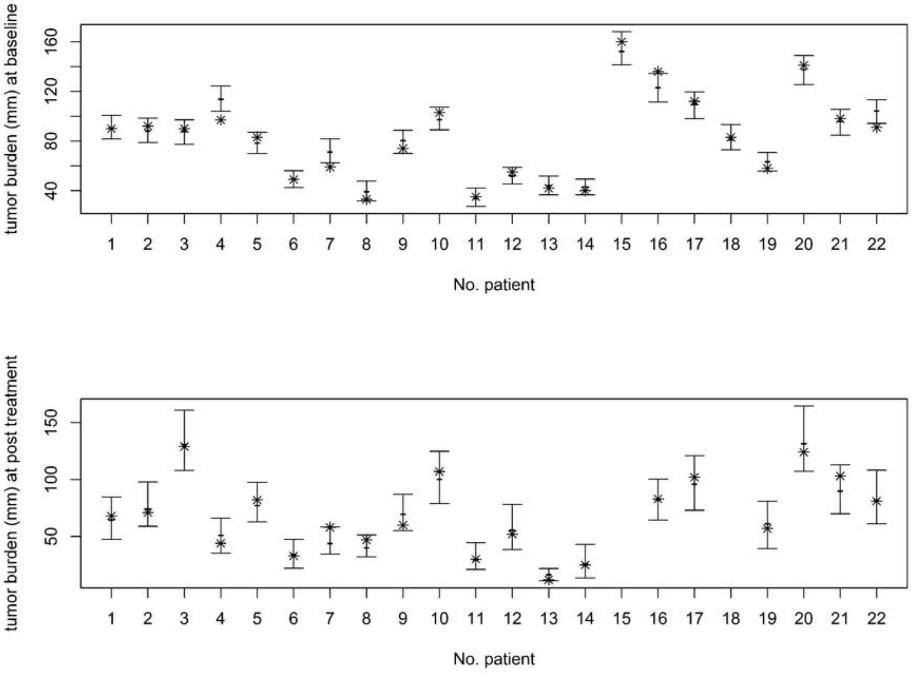
Results: The extent of differences between assessments of the percent change decreased non-linearly with the increase of baseline burden, and linearly with larger shrinkage of burdens. The probability for partial response or progression to result from reassessment had a sigmoid shape depending on the percent change in the first reading, inflecting at cutoff points (-30% and 20%, each). In three virtual trials having the same bas

trial 2. c. The probability of the diagnosis of partial response when reassessments was made by another reader for each patient in trial 3. (a) Distribution of the objective response rate in the reassessment of trial 1 with a 95% central range of (0.35, 0.65). (b) Distribution of the objective response rate in the reassessment of trial 2 with a 95% central range of (0.4, 0.6). (c) Distribution of the objective response rate in the reassessment of trial 3 with a 95% central range of (0.45, 0.65).

Validation of the model and the simulation process

Using an external dataset, the actual second measurement value was located within the 95% central range of hypothetical second readings in 19 of 22 (86%) and 20 of 20 (100%) burdens at baseline and post-treatment, respectively (Fig 4A). For the 20 resampled datasets, the ORR resulting from a real second assessment lay in the 95% central range of simulated ORRs in all the datasets (Fig 4B). Using a sponsor-initiated trial dataset with a provision of measurement values of burdens assessed by a BICR, the ORR based on the actual BICR lay within the range of the simulated ORRs in 19 of the 20 resampled datasets.

A. Tumor burden size per patient



B. ORR per resampling

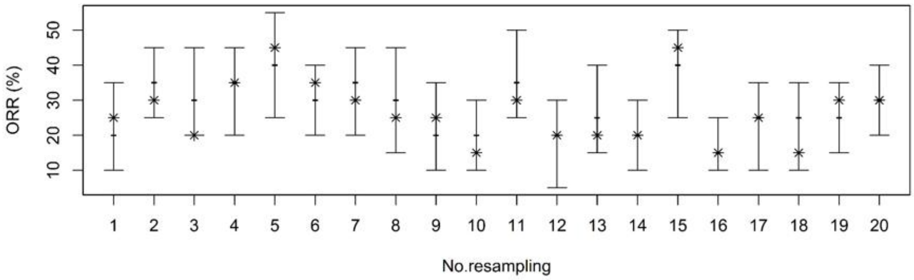


Figure 4. Interval plots of results for the validation and real measurements.

A: Tumor burden size per patient at each baseline (the above plot) and post-treatment (the below plot). The intervals indicate the 95% central ranges of the simulated burden sizes. The stars indicate the tumor burden sizes re-measured by the second radiologist.

B: ORR per resampling. The intervals indicate the 95% central ranges of overall response rates (ORRs) from the simulation. The circles and stars indicate the ORRs measured by the first radiologist and the second radiologist, respectively.

DISCUSSION

The extent of measurement variability between assessments of burdens was explored in some limited scenarios (6-9, 13). We investigated the extent of this phenomenon in various scenarios that are applicable to diverse burden sizes and percent changes in oncology trials. The extent of measurement variability was influenced by the baseline size of the tumor burden and the magnitude of its percent change after treatment. In the circumstances where smaller differences were obtained, a sharper slope was observed in the curve representing the probability that the percent change would be categorized as partial response (or progression) at the second assessment. A sharper slope indicated a more reproducible result in the designation of tumor response between assessments. These findings generally support the recommendations for the selection and measurement of target lesions in the RECIST guidelines version 1.1 (4). When clinicians make an important treatment decision in an individual oncology patient such as a change or withhold of chemotherapy regimen, they may refer the simulated probability curve to determine how an observed percent change is robust against the measurement variability.

The dataset used for the modeling in this study involved expert radiologists from independent affiliations. This situation could be considered as a simple form of BICR (14). When compared with a

BICR, our simulations provided a range of results representing 1000 independent reassessments. Considering the resources required for BICRs and the paucity of tools for evaluating the reproducibility of reported ORRs in practice, the adoption of our 95% central range would be a pragmatic alternative or adjunct to BICRs. More specifically, the 95% central range might potentially guide whether a phase 2 trial proceeds to a phase 3 trial by reviewing the robustness of the observed ORR against the measurement variability in a phase 2 trial. Nonetheless, it cannot entirely replace BICRs, because the simulation cannot detect systemic bias in the local review towards over-estimation or under-estimation of tumor shrinkage (15).

The 95% central range of the ORR should be interpreted differently from a 95% CI of the ORR. The 95% CI deals with the uncertainty against sampling errors in the estimation of the ORR (16), while the 95% central range of the ORR deals with the reproducibility of the ORR despite measurement variability. Current investigation could not simultaneously handle the variation of ORR which result from the sampling error and measurement variability. In oncology trials reporting the ORR, it would be useful to report the 95% central range of the ORR alongside the 95% CI for readers to be able to assess not only the degree of precision for the estimated ORR, but also its robustness against measurement variability.

We were not able to quantify the impact of measurement variability on PFS since measuring PFS involves not only the binary events of progression, but also the timing of the events (1). However, it would be at least possible to evaluate the reproducibility of progression, as a binary event, in terms of a progression rate. When patients experience unequivocal radiologic progression, symptomatic progression, or death, a probability of 1 could be assigned for their diagnosis with progression at the time of the repeated measurement, since those cases are not affected by measurement variability. However, unequivocal signs of radiologic progression on post-treatment CT scans, such as the appearance of a new target lesion or the progression of a non-target lesion, may be inconsistently perceived between assessments, meaning that assigning a fixed probability of 1 in such cases may over-estimate the reproducibility of the designation of progression.

Practically, measurable target lesions are selected to define the tumor burden in every assessment of baseline tumor response (4). However, in our study, the predefined target lesions were repeatedly measured in every assessment between readers or by a single reader, thus, our modeling could not reflect the measurement variability which originated from different target lesion selection. Different selections of target lesions could result in a greater difference in the percent change between assessments, and consequently could result in a less reproducible tumor response classification (17). Indeed, the proportion of concordantly selected target lesions was 41% in a group of three readers, and agreement rates of RECIST-based response classification among readers were much lower with discordant target lesion selection (18). Nevertheless, our simulation tool appropriately predicted the variation of ORR even in a setting where the target lesions were selected differently by the BICR (Fig 4 (B)). Further investigation of potential influence of lesion selection would be useful.

We initially performed subgroup analyses for the measurement variability according to the location, margin, and conglomeration of a target lesion. Since the margin and conglomeration substantially depended on the lesion size and the estimated distributions of measurement errors had the size dependency, the lesion size itself was sufficient to explain the measurement variability. The location of the lesion had little effect on the measurement variability. Thus, they were not considered in the final model.

The generalizability of our modeling and simulation tool may be an issue. Concerning the representativeness of the dataset used for our modeling, all 6 readers had at least 10 years of clinical experience in oncology body imaging, which is sufficient for representing the expert readers who participate in BICRs. With regard to sampling variability, we performed the simulation processes iteratively by changing the variance-covariance matrix of the distribution of measurement errors by re-sampling each component from its posterior distributions, presenting the extent of uncertainty as noted with bands in Fig 1 and lines in Fig 2.

In conclusion, determinations of partial response or progression in oncology trials are probabilistic outcomes that follow a certain probability distribution due to measurement variability. Quantification of how much a given result could be changed by accounting for measurement variability would help inform any further decisions made on the basis of trial result. Further study is required to extend this work to incorporate the quantification of the reproducibility of PFS against measurement variability.

REFERENCES

1. Wilson MK, Karakasis K, Oza AM. Outcomes and endpoints in trials of cancer treatment: the past, present, and future. The Lancet Oncology. 2015;16(1):e32-42.

2. Johnson KR, Ringland C, Stokes BJ, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. The Lancet Oncology. 2006;7(9):741-6.

3. Prasad V, Kim C, Burotto M, et al. The Strength of Association Between Surrogate End Points and Survival in Oncology A Systematic Review of Trial-Level Meta-analyses. JAMA Intern Med. 2015;175(8):1389-98.

4. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). European Journal of Cancer. 2009;45(2):228-47.

5. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. Journal of Clinical Oncology. 2003;21(13):2574-82.

6. Oxnard GR, Zhao B, Sima CS, et al. Variability of lung tumor measurements on repeat computed tomography scans taken within 15 minutes. Journal of Clinical Oncology. 2011;29(23):3114-9.

7. Zhao B, James LP, Moskowitz CS, et al. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. Radiology. 2009;252(1):263-72.

8. Yoon SH, Kim KW, Goo JM, et al. Observer variability in RECIST-based tumour burden measurements: a meta-analysis. European Journal of Cancer. 2016;53:5-15.

9. Shao T, Wang L, Templeton AJ, et al. Use and misuse of waterfall plots. Journal of the National Cancer Institute. 2014;106(12). pii: dju331.

10. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 1986;327(8476):307-10.

11. Perini TA, Oliveira GLd, Ornellas JdS, et al. Technical error of measurement in anthropometry. Revista Brasileira de Medicina do Esporte. 2005;11(1):81-5.

12. Euser AM, Dekker FW, le Cessie S. A practical approach to Bland-Altman plots and variation coefficients for log transformed variables. Journal of Clinical Epidemiology. 2008;61(10):978-82.

13. McErlean A, Panicek DM, Zabor EC, et al. Intra- and interobserver variability in CT measurements in oncology. Radiology. 2013;269(2):451-9.

14. U.S. Food and Drug Administration. Guidance for industry:

developing medical imaging drug and biological products, part 3: design, analysis, and interpretation of clinical studies. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071604.pdf

15. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: Important design element or unnecessary expense? Journal of Clinical Oncology. 2008;26(22):3791-6.

16. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. British Medical Journal. 1986;292(6522):746-50.

17. Keil S, Barabasch A, Dirrichs T, et al. Target lesion selection: an important factor causing variability of response classification in the Response Evaluation Criteria for Solid Tumors 1.1. Investigative radiology. 2014;49(8):509-17.

18. Kuhl CK, Alparslan Y, Schmoee J, et al. Validity of RECIST Version 1.1 for Response Assessment in Metastatic Cancer: A Prospective, Multireader Study. Radiology. 2018 Nov 6:180648. doi: 10.1148/radiol.2018180648.

국문 초록

목적: 종양 총량 평가간 측정 변이를 탐색하고, 종양 임상시험에 적용이 가능하도록 치료반응 분류 및 객관적 반응률에 대한 측정 변이의 영향을 탐색하고자 함.

방법: 임상시험 전산화 단층촬영 데이터를 이용하여 측정 변이에 대한 계층적 모델은 만들고 모델을 이용하여 다음의 목적을 위하여 일련의 시뮬레이션을 시행하였다. 1) 종양 총량의 변화율을 반복 평가했을 때 측정 변이가 어느 정도 발생하는지 파악함. 2) 측정 변이에 의해 치료 반응 분류가 확률적으로 어떻게 변하는지를 실증함. 3) 재평가했을 때 측정 변이에 의한 객관적 반응률의 변동을 95% 중앙 범위를 생성함으로써 추정함.

결과: 종양 총량 변화율을 반복 평가했을 때 종양 총량 변화율에 대한 측정 변이의 범위는 기저 종양 총량이 증가할수록 비선형적으로 감소하고 치료 후 총량이 감소할수록 선형적으로 감소하였다. 재평가 했을 때 부분 반응 또는 진행으로 평가될 확률은 첫 번째 측정된 변화율에 의존하여 sigmoid 분포를 보이며 부분 반응 기준점 변화율인 -30% (또는 진행의 경우 20%)에서 변곡점을 보였다. 동일한

기저 종양 총량 및 동일한 50% 객관적 반응률을 보이는 3 가지 가상

임상시험에서 부분 반응 기준점인 −30% 주위에 종양 총량 변화율들이 적게 분포할수록 객관적 반응률이 재현될 확률이 보다 높았다 (35%−65% vs. 40%−60% vs. 45%−60%).

결론: 부분 반응 및 진행 판정은 측정 변이에 영향을 받는 확률적 결과이므로 잠재적 측정변이에 의한 객관적 반응률의 변동을 정량화 하는 것이 필수적이며, 임상시험 결과에 기반한 의사 결정에 도움이 될 것으로 기대된다.

주요어 : 종양, 반응, 측정, 변이, 객관적 반응률
학　번 : 2008−23269