Editorial

Potential benefits of large database analysis

Slav Yartsev

London Regional Cancer Program, London Health Sciences Centre, London, ON, Canada

Correspondence to: Slav Yartsev. London Regional Cancer Program, London Health Sciences Centre, London, ON, Canada.

Email: Slav.Yartsev@lhsc.on.ca.

Provenance: This is a Guest Editorial commissioned by Section Editor Hongcheng Zhu, MD, PhD (Department of Radiation Oncology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China).

Comment on: Lin SH, Zhang N, Godby J, *et al.* Radiation modality use and cardiopulmonary mortality risk in elderly patients with esophageal cancer. Cancer 2016;122:917-28.

Submitted Jul 13, 2017. Accepted for publication Jul 24, 2017.

doi: 10.21037/atm.2017.08.02

View this article at: http://dx.doi.org/ 10.21037/atm.2017.08.02

The development of sophisticated technologies for radiation therapy and new targeting drugs in chemotherapy presents various options for cancer treatment. These new options may be quite costly and this aspect should be taken into account due to health care burden to the society worldwide (1). Every time the treating physician faces the problem of the best/optimal way to treat a new patient. Where can he or she find help? Clinical trials are sometimes contradictory and prospective analyses provide little evidence due to limited patient cohort. In recent *Annals for Translation Medicine* editorial (2), Salem *et al.* argued that the nonrandomized, retrospective observational analysis by van Diepen *et al.* (3) generated more questions than answers and proposed an adequately powered, prospective randomized clinical outcome trial as the proper solution. But imagine how costly would be such trial and how long will it take for patient accrual and collection of the outcomes data. There is also a strong possibility that, when the results are finally published, a new, even more promising, solution will be put on the market.

Two clinically important questions: 3D conformal radiotherapy (3DCRT) *vs.* intensity-modulated radiation therapy (IMRT) and the factors to correlate with cardiopulmonary mortality risk were explored by Lin *et al.* in a manuscript published by *Cancer* (4). The authors analyzed the outcomes for 2,240 patients with esophageal cancer treated with 3DCRT and 313 patients treated with IMRT between 2002 and 2009. The patient demographic information included age, sex, race/ethnicity, marital status, geographical location, urban/rural setting, education, income level. Tumor characteristics included stage, grade, and year of diagnosis. Treatment characteristics included the use of chemotherapy within 6 months of diagnosis and esophagectomy after radiotherapy. Interestingly, the authors also collected information regarding the physician's age, sex, primary and secondary specialties, board certification, US trained or not, number of years in practice after training, and esophageal cancer case load. Almost all included patient demographic, clinical and tumor characteristics were the same for 3DCRT and IMRT patient cohorts, except obvious lower proportion of IMRT case at earlier years of diagnosis when IMRT was only gaining popularity. The authors used inverse probability of treatment weighting (IPTW) Cox model analysis to compare the outcomes of the two groups. IMRT was found significantly associated with lower all-cause mortality and cardiac-specific mortality, but not cancer-specific and pulmonary mortality compared to 3DCRT treatments. Inclusion of physician's data showed that younger physicians used IMRT significantly more frequently than older physicians. The latter finding points to the necessity of better dissemination of the latest technological advances among more seasoned physicians. This population-based analysis of Lin *et al.* provided evidence that IMRT should be preferred choice of treatment of patients with esophageal cancer without conducting costly prospective clinical trials.

Vickress *et al.* created a multivariable model for prediction of survival for liver cancer patients treated with radiation therapy (5). This study is a good example of additional knowledge that may be obtained from large retrospective studies. The authors have included both pre-treatment patient information and parameters of radiation treatment itself and constructed two practical nomograms for primary hepatocellular carcinoma and liver metastasis patients. They also found that the volume of normal liver receiving more than 24 Gy (V24) correlated positively with disease free survival for rotational modalities of radiation delivery [volumetric-modulated arc therapy (VMAT) and TomoTherapy], but not for the patients treated with fixed-beams IMRT. This finding indicates that relatively large doses delivered in the area adjacent to the tumor in an isotropic manner are able to eradicate microscopic spread of malignant cells, while fixed-beam delivery misses these cells in the regions between the beams. In a letter to *Future Oncology* Vickress *et al.* hypothesized that the density of neoplastic cells was decreasing gradually with distance from the gross tumor (6). The lower is the density of malignant cells, the better is their oxygenation. Far from the tumor the patient's immune system or targeted chemotherapy (7) can eliminate residual disease, but one may need to add radiation to eradicate cancer spread in the area closer to the tumor to result in a disease-free survival advantage. Vickress *et al.* proposed that the planned dose should gradually decrease around the tumor in order to match the diminishing density of neoplastic cells. This finding is intriguing, even though it may be a spurious statistical anomaly, warranting validation in other databases. If this hypothesis can be validated, it would prompt a reversal of current trends in management philosophy.

Both Lie *et al.* and Vickress *et al.* studies show that a retrospectives analysis of the treatment outcomes even for a patient cohort of intermediate size may be useful in establishing valuable information. However, they had insufficient statistical power to give definite recommendations for a particular patient. Michael Porter from the Harvard Business School has initiated a creation of the International Consortium for Health Outcomes Measurements (ICHOM) that aims at providing a global resource of in-use outcome measures and risk adjustment factors by medical condition (8). It is important to establish what kind of information should be collected in such databases in order to make it useful for creation of mathematical model that will allow to predict a probability of different treatment scenarios for cohorts of patients based on their patient specific pre- and post-treatment information leading to personalized patient treatment as proposed by Yartsev and Mackie (9). A creation of the database that includes patient specific information before treatment, planning parameters, actual radiation delivery, and detailed follow-up information of acute and late toxicities, quality of life, and patient satisfaction can allow for an informed choice of the treatment and planning procedures for a new patient. Ulfenborg *et al.* proposed the five-stage (data preparation, exploratory data analysis, confirmatory analysis, biological knowledge discovery, and visualization of the results) data analysis framework for biomedical big data (10) that can be applied for other databases as well. Obvious solution would be a coordinated data collection by many cancer centers and analysis. In this process data sharing and data merging raise legal, policy, and technical issues need to be addressed (11). The realization of data sharing network requires not only tools to allow interconnection and global integration of all clinical data but also a universal legal framework to protect the privacy of every patient. Embracing new technologies in data collection and analysis combined with active participation of multiple cancer centers can be a reliable and cost-effective way to personalized patient care.

Acknowledgements

None.

Footnote

Conflicts of Interest: The author has no conflicts of interest to declare.

References

1. Defourny N, Dunscombe P, Perrier L, et al. Cost evaluations of radiotherapy: What do we know? An ESTRO-HERO analysis. Radiother Oncol 2016;121:468-74.
2. Salem S, Askandar S, Khouzam RN. Aspirin monotherapy vs. dual antiplatelet therapy in diabetic patients following coronary artery bypass graft (CABG): where do we stand? Ann Transl Med 2017;5:213.
3. van Diepen S, Fuster V, Verma S, et al. Dual antiplatelet therapy versus aspirin monotherapy in diabetics with multivessel disease undergoing CABG: FREEDOM insights. J Am Coll Cardiol 2017;69:119-27.
4. Lin SH, Zhang N, Godby J, et al. Radiation modality use and cardiopulmonary mortality risk in elderly patients with esophageal cancer. Cancer 2016;122:917-28.
5. Vickress J, Lock M, Lo S, et al. A multivariable model

to predict survival for patients with hepatic carcinoma or liver metastasis receiving radiotherapy. Future Oncol 2017;13:19-30.
6. Vickress J, Lock M, Lo S, et al. Potential benefit of rotational radiation therapy. Future Oncol 2017;13:873-4.
7. Qian Y, Zeng ZC, Ji Y, et al. Microinvasion of liver metastases from colorectal cancer: predictive factors and application for determining clinical target volume. Radiat Oncol 2015;10:125.
8. Available online: http://ichom.org
9. Yartsev S, Mackie TR. Comment on "When is better best? A multiobjective perspective" [Med. Phys. 38, 1635-1640 (2011)]. Med Phys 2013;40:077101.
10. Ulfenborg B, Karlsson A, Riveiro M, et al. A data analysis framework for biomedical big data: Application on mesoderm differentiation of human pluripotent stem cells. PLoS One 2017;12:e0179613.
11. Gal J, Milano G, Ferrero JM, et al. Optimizing drug development in oncology by clinical trial simulation: Why and how? Brief Bioinform 2017. [Epub ahead of print].

Cite this article as: Yartsev S. Potential benefits of large database analysis. Ann Transl Med 2017;5(20):413. doi: 10.21037/atm.2017.08.02