

ビッグデータとモデルの関係性について

北海道大学情報基盤センター 水田正弘

1. はじめに

2011 年前後から注目され始めた「ビッグデータ」は、実社会へ大きな影響を与えるまでに普遍化した。医療をはじめとする分野において大量のデータが有するポテンシャルについては疑問の余地はないが、その活用法については多くの検討課題がある。解釈可能なモデルを仮定しないいわゆるブラックボックスモデルも、単に結論を出すためには有効である。しかし、結論の妥当性や適用範囲を評価するには不十分である。数理モデルまたは統計モデルを設定し、より優れた知見を見出す方法が考えられる。

本報告では、放射線治療に関する数理モデルを例にとり、このような問題について報告する。

2. ホワイトボックスモデルとブラックボックスモデル

データの解析で想定するモデルとして、解釈が可能なホワイトボックスモデルと解釈が困難なブラックボックスモデルがある。統計学専門家として、後者の欠点だけを強調することは容易であるが、実際にデータと対峙している人に対しては、説得力がない。データ過多時代におけるデータとモデルを真摯に考えなくてはいけない。

データ主導は重要な観点である。しかし、適切なモデルを設定せずに、結論の妥当性や適用範囲を評価することはできない。ホワイトボックスモデルを構築し、そのパラメータ推定や検証にいわゆる「ビッグデータ」を用いることは有益であると考えられる。

3. 放射線治療における数理モデル

腫瘍に対する治療法として放射線治療がある。例えば、陽子線治療では従来の X 線による治療に比べて、腫瘍以外の正常組織（危険臓器）への被ばくを減少させることができる。そのことのみで陽子線治療の優位性が証明されたとは言い難い。本来 RCT により検証すべきであるが、現実問題としてその実施は困難である。

放射線に対する腫瘍を含む細胞の生存率を LQ モデルで表現することがある。これには 2 つのパラメータがある。報告者は、LQ モデルを仮定した場合の最適な放射線治療の照射回数および線量の導出法を報告している [1] [2]。その計画の妥当性を証明するためには、モデルの妥当性・頑健性・パラメータを検討しなくてはならない。

参考文献

[1] M. Mizuta, S. Takao, H. Date, N. Kishimoto, K.L. Sutherland, R. Onimaru, H. Shirato (2012) A Mathematical Study to Select Fractionation Regimen based on Physical Dose Distribution and the Linear-Quadratic Model, *International Journal of Radiation Oncology, Biology, Physics*, Vol.84, No.3, pp.829-833, doi:10.1016/j.ijrobp.2012.01.004

[2] Y. Sugano, M. Mizuta, S. Takao, H. Shirato, K. L. Sutherland, H. Date (2015) Optimization of the fractionated irradiation scheme considering physical doses to tumor and organ at risk based on dose volume histograms. *Medical Physics* vol.42 pp.6203-6210 doi: 10.1118/1.4931969