Published in final edited form as:

Urology. 2010 March ; 75(3): 511–513. doi:10.1016/j.urology.2009.07.1265.

Why can't Nomograms be more like Netflix?

Andrew J. Vickers, Paul Fearn, Mike W. Kattan, and Peter T. Scardino

Memorial Sloan-Kettering Cancer Center (AJV, PF and PTS) and Cleveland Clinic (MWK).

Two types of prediction: nomograms and Netflix

Nomograms have become ubiquitous in urology. As evidence: a simple PubMed search for “nomogram prostate” obtains over 450 hits; websites abound where patients can type in data and obtain predictions (including both www.nomogram.org and www.nomograms.org); nomograms have been used as inclusion criteria for clinical trials; even patient medical records are now designed so that nomograms can be calculated automatically from a patient's data. For members of the general public, however, the prediction model they are most likely to encounter is not a nomogram, but Netflix.

Netflix is a company that rents DVDs to paying subscribers through the mail. Subscribers choose which movies to rent via a web interface that allows browsing by genre and searches for specific films, actors or directors. Subscribers can also rate movies that they have seen on a five point scale, and then look up the average of all ratings for any movie they might consider renting. For example, a subscriber could look up on Netflix that, as of April 2009, *All about Eve* has an average of 3.9 stars whereas *Best in Show* has an average of 3.8 stars. But the real key of Netflix is its recommender system: a subscriber can look up not only the average rating, but a weighted average considering how that subscriber has rated other movies. For example, when the first author of this paper (AV) looks up *Best in Show* he is given “Average of raters like you: 4.1 stars”; for *All about Eve*, the average for raters like AV is 3.3, reversing the rank order suggested by the average ratings.

The Netflix recommender system is based on a computer-intensive algorithm. To explain this in simple terms, when AV looks up the movie *Best in Show*, the algorithm searches through its entire database of subscribers to find those who had rated this movie. It then looks at each subscriber in this subset to see if they had rated movies that AV had also rated. Based on how similar their ratings were to his, each subscriber's rating of *Best in Show* is given a weight, and then all ratings combined in a weighted average. This process is actually a pretty good reflection of how we normally filter advice; if a friend tells you he loved some new movie, you might think “but he also loved that one with Julia Roberts, which I thought was a bore” and take his advice with a pinch of salt; alternatively, you might think: “he and I tend to like the same sort of movies, so I'll probably enjoy this one”.

© 2009 Elsevier Inc. All rights reserved.

Corresponding author: Andrew J. Vickers Associate Attending Research Methodologist Department of Epidemiology and Biostatistics Memorial Sloan Kettering Cancer Center 307 East 63rd Street NY, NY 10021 Tel: 646 735 8142 Fax: 646 735 0011 Cell: 347 244 6934 vickersa@mskcc.org <http://www.mskcc.org/mskcc/html/3197.cfm>.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Why is Netflix preferable to nomograms?

A Netflix-type system has four major advantages to most nomograms.

1. Missing data

Calculating a predicted probability from a nomogram requires data on all predictors; if data on even one predictor are missing, no prediction is possible. Missing data are no problem for the Netflix algorithm, indeed missing data is ubiquitous: most subscribers have rated only a tiny fraction of the many thousands of movies on the site. What the algorithm does is to search for similar raters in terms of whatever information it has to hand. Take the prediction of recurrence after radical prostatectomy. If a patient had missing data on PSA, a Netflix-type system would define “similarity” only in terms of stage and grade.

2. Changing predictions with changing data

Nomograms are fixed and unchanging. This is a problem because both patients and treatments change over time. For example, the Kattan pre-operative nomogram for predicting recurrence after radical prostatectomy include patients treated in the early 1980's¹. The intervening period has seen considerable stage shift, dramatic changes in surgical caseloads and experience, modifications to Gleason grading and the development of laparoscopic and robotic surgery. For example, a Gleason 7 in 1991 is not the same as a Gleason 7 in 2009^{2, 3}; similarly, it is difficult to be confident a patient treated by an open surgeon on the early part of the learning curve⁴ had a similar chance of cure as when treated by a contemporary high-volume surgeon using a laparoscopic approach⁵. In contrast, Netflix predictions are updated in real time as information becomes available: the *Hannah Montana Movie*, for example, is popular now, but will likely seem dated in 20 years time. The Netflix rating will evolve over time and reflect this change in taste.

3. Changing predictions with changing status

According to one recent nomogram⁶, a patient with a Gleason 7, organ-confined tumor, PSA of 6 and negative margins, has a 10% probability of recurrence at 2 years. The problem with this prediction is that it is static. If the patient has a PSA of zero 18 months after surgery, his chance of recurrence at two-years no doubt falls below 10%, yet the nomogram spits out the same prediction. Netflix, on the other hand, can update predictions in real-time. As a simple example, AV's predicted rating for *Friday the 13th Part 8: Jason takes Manhattan* is 2.2 stars. However, if he rented a bunch of horror movies and gave them all high ratings, the Netflix algorithm would adjust and increase his predicted rating for this movie.

4. Flexibility

Netflix can provide a predicted rating for absolutely any movie in its database. In contrast, each nomogram gives only a specific prediction. Indeed, the prevailing wisdom seems to be that a new nomogram is required for each possible combination of predictors, for each outcome (e.g. recurrence vs. cancer specific death) and treatment (e.g. surgery and radiotherapy) at each follow-up time. So, for example, we now have nomograms to predict recurrence based on both pre-operative¹ and postoperative⁷ data; recurrence at 2 years⁶, recurrence at 5 years incorporating transforming growth factor-beta¹⁸, recurrence at 5 years including surgeon experience in the model⁹, the trifecta of continence, potency and freedom from recurrence¹⁰, prostate cancer specific death at 10 years¹¹; indolent cancer¹² and so on and so forth. Indeed, a review to July 2007 found over 100 different prediction tools for prostate cancer¹³. It can be estimated that, given a set of predictor variables, some of which may be missing, different endpoints (pathology, recurrence, death, continence, potency), and predictions made for and at various times after surgery or radiotherapy, the total number of nomograms needed would

approach the 6,000 papers published worldwide on prostate cancer each year. So if we dedicated, say, 25% of the entire world literature only to nomograms, it would take 4 years to publish them all. At which point, of course, things would have changed, new data would be available and we would have to start all over again.

Could we have a Netflix for medicine?

Nomograms are already starting to look a little bit like Netflix. The best example is the prostate cancer prediction tools available at the www.nomograms.org website. These can deal with missing data, up to a point, by making available different nomograms depending on what data are available; some of the nomograms can also be updated for changing patient status, for example, by modifying a prediction about recurrence if the patient is recurrence-free at one year. There have also been systematic attempts to define the range of different nomograms that are needed (the “metagram” concept)¹⁴.

But could we go beyond current technologies that and have a full Netflix-type system? To show how this might work, we will use, as an example, a system to predict outcome after surgery for prostate cancer. This would be based on a large dataset of baseline characteristics and outcome data from radical prostatectomy patients. Take the case of a patient, Mr. Brown, with data on age, preoperative erectile function, PSA and pathologic stage and grade. Mr. Brown is 3 months out, has poor erectile function and wants to know whether this means he'll likely have long-term impotence. The database would first search for patients with data on erectile function say, at least two years after surgery. It would then use those patients to create a prediction for Mr. Brown. The exact methods are perhaps not particularly important. In theory, logistic regression could be used, though in practice, computationally intensive methods are more common because of their flexibility; for example, such methods could use data from patients in even if they had missing data for one of Brown's predictors. If Mr. Brown himself was part of the data set, his prediction could be documented and then compared with his erectile function status at two years. In this way, the prediction system would continually be evaluating its own performance in real time.

All this begs the question of how a large dataset of radical prostatectomy outcomes might be made available. We have already had some ideas on that front: many US centers use a standard database system, known as Caisis, for management of prostate cancer patients and these different databases could be linked to create a single, master data set.

Conclusion

A nomogram is defined as a graphical display of a prediction model or calculation tool. In the traditional implementation of the nomogram, the user draws pencil lines between axes, counts up points, and then reads off a prediction. The development of modern information technology allows automatic linkage of geographically disparate data sets with dynamically updated predictions in real time. Nomograms may well have changed the way that many clinicians and researchers think about medicine, but it is time to accept that there are technologically superior solutions to the problem of medical prediction.

References

1. Stephenson AJ, Scardino PT, Eastham JA, et al. Preoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Natl Cancer Inst* 2006;98:715. [PubMed: 16705126]
2. Egevad L. Recent trends in Gleason grading of prostate cancer: I. Pattern interpretation. *Anal Quant Cytol Histol* 2008;30:190. [PubMed: 18773736]

3. Epstein JI, Allsbrook WC Jr, Amin MB, et al. Update on the Gleason grading system for prostate cancer: results of an international consensus conference of urologic pathologists. *Adv Anat Pathol* 2006;13:57. [PubMed: 16462155]
4. Vickers AJ, Bianco FJ, Serio AM, et al. The surgical learning curve for prostate cancer control after radical prostatectomy. *J Natl Cancer Inst* 2007;99:1171. [PubMed: 17652279]
5. Vickers AJ, Savage CJ, Hruza M, et al. The surgical learning curve for laparoscopic radical prostatectomy: a retrospective cohort study. *Lancet Oncol* 2009;10:475. [PubMed: 19342300]
6. Walz J, Chun FK, Klein EA, et al. Nomogram predicting the probability of early recurrence after radical prostatectomy for prostate cancer. *J Urol* 2009;181:601. [PubMed: 19084864]
7. Stephenson AJ, Scardino PT, Eastham JA, et al. Postoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Clin Oncol* 2005;23:7005. [PubMed: 16192588]
8. Shariat SF, Karam JA, Walz J, et al. Improved prediction of disease relapse after radical prostatectomy through a panel of preoperative blood-based biomarkers. *Clin Cancer Res* 2008;14:3785. [PubMed: 18559597]
9. Kattan MW, Vickers AJ, Yu C, et al. Preoperative and postoperative nomograms incorporating surgeon experience for clinically localized prostate cancer. *Cancer* 2009;115:1005. [PubMed: 19156928]
10. Eastham JA, Scardino PT, Kattan MW. Predicting an optimal outcome after radical prostatectomy: the trifecta nomogram. *J Urol* 2008;179:2207. [PubMed: 18423693]
11. D'Amico AV, Moul J, Carroll PR, et al. Cancer-specific mortality after surgery or radiation for patients with clinically localized prostate cancer managed during the prostate-specific antigen era. *J Clin Oncol* 2003;21:2163. [PubMed: 12775742]
12. Dong F, Kattan MW, Steyerberg EW, et al. Validation of pretreatment nomograms for predicting indolent prostate cancer: efficacy in contemporary urological practice. *J Urol* 2008;180:150. [PubMed: 18485398]
13. Shariat SF, Karakiewicz PI, Roehrborn CG, et al. An updated catalog of prostate cancer predictive tools. *Cancer* 2008;113:3075. [PubMed: 18823041]
14. Nguyen C, Kattan M. Development of a prostate cancer metagram: a solution to the dilemma of which prediction tool to use in patient counseling. *Cancer*. 2009 In Press.