Comment on POMDP Approach to Personalize Mammography Screening Decisions

Diana Petitti, MD, MPH

Department of Biomedical Informatics
Center for Health Information and Research
Arizona State University
502 E. Monroe Street, Suite C320
Phoenix, Arizona
602-496-2015

Diana.petitti@asu.edu

Recognition that the amount of benefit of a screening test is determined by the probability of the disease in the people being screened is longstanding. For cancer and other chronic diseases, the probability of the disease in a previously unscreened population is defined by the disease prevalence. In a previously screened population, the probability of the disease is defined by disease incidence. The attempt to tailor recommendations about screening based on a person's underlying probability of disease using information on epidemiologic and other risk factors has a long history.

Thus, the recommendation by a number of expert bodies to begin screening for colorectal cancer at age 50 is based on the epidemiology of colorectal cancer, which shows a fairly sharp increase in the prevalence of colorectal cancer beginning at age 50 years. Recommendations by a number of expert bodies to begin to screen people with a family history of colorectal cancer at an age less than 50 years is based on the epidemiologic evidence showing the prevalence of colorectal cancer is higher at a younger age in people with a family history of colorectal cancer. Recommendations by a number of expert bodies to rescreen people with a negative screening colonoscopy every ten years--not every year or every five years--is based on evidence that shows that the incidence of colorectal cancer after a negative screening test is low.

Risk-stratification based on disease probability is a fundamental principle of screening. Use of single risk factors, such as age or family history, to stratify people according to their risk has been increasingly supplanted by models that incorporate multiple risk factors to stratify people according to their risk.

That screening tests have harms as well as benefits is also well-recognized. Consideration of the net benefit of screening when making recommendations to individuals about whether to be screened and in making policies meant to apply broadly is critically important. Models permit simultaneous estimation of risks and benefits. Models permit things like time-to-benefit and competing risks to be considered when estimating benefits and risks. As pointed out by Ayer, Alagoz and Stout, use of models could improve the efficiency of screening programs and reduce the burden of screening for individuals and society.

Mitchell Gail and colleagues at the National Cancer Institute were the first to use risk factors for breast cancer to develop a formal model to predict disease probability, in specific the chances of

developing breast cancer over a specific time horizon (Gail 1989). The so-called "Gail model" was used to identify women at high risk of breast cancer eligible to enter a randomized trial of tamoxifen versus placebo (Fisher 1998) and a separate trial comparing tamoxifen and raloxifene (Vogel 2006) for the primary prevention of breast cancer. The potential use of this model to make risk-based recommendations about mammography screening for women in their forties was described in 1998 (Gail and Rimer 1998). Other models that might be useful in tailoring mammography screening recommendations exist and have been promoted specifically for use to tailor mammography screening (Mandelblatt 1992; Tice 2008; Schousboe 2011).

The partially observable Markov decision process (POMDP) model described by Ayer, Alagoz and Stout is another effort at developing a model that could be used to tailor recommendations about screening mammography. It is an especially well-crafted extension of prior modeling efforts aimed at providing better information upon which individual and "personalized" recommendations about mammography screening. The dynamic nature of the model is a noteworthy advance. By incorporating prior screening history and outcomes of screening, the model permits identification of women who do not need to be screened (or need to be screened less often) because they are a low risk of developing breast cancer. The incorporation of information on the age-dependence of disease progression, mortality, and test accuracy is also important.

I would not hesitate to use the model to decide whether I (or a friend or a relative or a patient) should or should not have a screening mammogram. But my personal enthusiasm for the use of information from models to guide decisions about mammography screening is tempered by the painful recognition that my enthusiastic embrace of numbers for decision-making may be aberrant. My expectation that model-based information would be embraced as useful proved to be naively out of touch. This recognition arises from experiences and observations made during and after the release of the updated 2009 United States Preventive Services Task Force (USPSTF) mammography screening recommendations (USPSTF 2009). I interacted extensively with organizations, the media, women, and physicians in efforts to explain the recommendations and the models that provided information used to make the recommendations in the period after they were released. I have followed carefully the results.

The 2009 recommendations were an update of 2002 USPSTF recommendations on the same topic. There were two important changes in the mammography screening recommendations. One related to screening of women at ages less than 50 and one related to screening interval. With regard to screening interval, the 2002 USPSTF recommended screening women age "every 1-2 years" whereas biennial screening was recommended in 2009. With regard to age, the 2002 USPTSF recommendation was to screen women age 40+ years. The 2009 recommendation was to screen women between 50 and 74 years routinely but to make an individualized decision to start regular screening before the age of 50 taking into account patient context, including values regarding specific benefits and harms.

The 2009 recommendations were strongly influenced by information from modeling work done by the Cancer Intervention and Surveillance Modeling Network (CISNET) Collaborators that had been commissioned by the USPSTF (Mandelblatt 2009) with the goal of making it possible to make more precise--that is tailored--recommendations about age to start to screen and

screening interval. The model results indicated that the net benefits of starting to screen at age 40 compared with age 50 and screening annually compared with biennially were quite small.

The mammography screening recommendations received instantaneous and almost uniformly negative attention from the print media, with front-page articles in the New York Times, the Washington Post, USA Today, among others. The recommendations were the subject of broadcasts on National Public Radio, CNN, the McNeill Lehrer report, the Today Show, and Good Morning America. Most of this coverage involved interviews that media professionals would identify as "hostile." On December 2, 2009, the House Subcommittee on Health of Energy and Commerce Committee (which oversees legislation related to health) held hearings at which the USPSTF Chair and Vice-Chair (me) underwent several grueling hours of questions by members of the committee about the motives of those making the recommendations and questioning (without evidence) the conclusions. The United States Congress ultimately negated any impact of the recommendations on payment by including in the Affordable Care Act language that stated that all recent recommendations of grade A or B made by the USPSTF must be paid for in any commercial insurance package but that the 2002 USPSTF mammography screening recommendations, not the 2009 recommendations, were the most recent.

My overall conclusions about models to guide recommendations, decisions and policy about mammography and to "tailor" or individualize these recommendations based on models flowing from this experience are as follows. First, there is essentially no interest of women and their physicians in numeric data from models as a guide to decision making about mammography, unless perhaps the data were to support a recommendation of "mammography every day, for every woman, forever." Second, the ability of women, physicians, and the media to understand information from models is so limited as to make potentially useful model essentially useless. Third, with only a few exceptions, policymakers in the United States are unwilling to use models as input to decisions that would mean that something as "popular" as mammography would be unavailable to anyone, no matter how small the benefit.

The lack of impact of the model-based recommendations of the USPSTF with regard to screening mammography is depressingly well-documented (Squiers 2011; Yasmeen 2012). Women ignore them and physicians hate them.

Can a better model turn the tide of mammography overscreening? Would an average, even an above average physician, be able to use a better model better? Given the sorry state of numeracy of the general population and the inherent complexity of the models that accurately capture the complexity of the underlying issue of tailored screening, can modelers hope to convince women, physicians, the media and policymakers of their utility? Depressingly, for now, I conclude that the answer to all of these questions is no.

Congratulations to Ayer, Alagoz, and Stout notwithstanding. Keep up the fight. But buy armor or develop a thick skin. On reflection, do both.

References

Fisher, B., J.P. Costantino, D.L. Wickerham, et al. 1998. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *Journal of the National Cancer Institute* **90**(18):1371-1388.

Gail, M., M. Rimer. 1998. Risk-based recommendations for mammographic screening for women in their forties. *Journal of Clinical Oncology* **16**(9):3105-3114.

Gail, M.H., L.A. Brinton, D.P. Byar, et al. 1989. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute* **81**(24):1879-1886.

Mandelblatt, J.S., K.A. Cronin, S. Bailey, et al. 2009. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Annals of Internal Medicine* **151**(10):738-747.

Mandelblatt, J.S., M.E. Wheat, M. Monane, et al. 1992. Breast cancer screening for elderly women with and without comorbid conditions. A decision analysis model. *Annals of Internal Medicine* **116**(9):722-730.

Schousboe, J.T., K. Kerlikowske, A. Loh, S.R. Cummings. 2011. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Annals of Internal Medicine* **155**(1):10-20.

Squiers, L.B., D.J. Holden, S.E. Dolina, et al. 2011. The public's response to the U.S. Preventive Services Task Force's 2009 recommendations on mammography screening. *American Journal of Preventive Medicine* **40**(5):497-504.

Tice, J.A., S.R. Cummings, R. Smith-Bindman, et al. 2008. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Annals of Internal Medicine* **148**(5):337-347.

US Preventive Services Task Force. 2009. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* **151**(10):716-726.

Vogel, V.G., J.P. Costantino, D.L. Wickerham, et al. 2006. Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: the NSABP Study of Tamoxifen and Raloxifene (STAR) P-2 trial. *Journal of the American Medical Association* **295**(23):2727-2741.

Yasmeen, S., P.S. Romano, D.J. Tancredi. 2012. Screening mammography beliefs and recommendations: a web-based survey of primary care physicians. *BMC Health Services Research* **12**:32.