Annual Combined Mammography and Tomosynthesis Screening: Is It Really Cost-Effective?

Christoph I. Lee[1,2,3]
Janie M. Lee[1,3]
Anna N. A. Tosteson[4]

OBJECTIVE. Examining the comparative effectiveness of adding tomosynthesis to mammography screening in the United States would help key stakeholders, including physicians and policy makers, guide shared decision-making and recommendations for routine breast cancer screening.

CONCLUSION. More robust data from U.S.-based longitudinal observational studies, clinical trials, or both are needed, along with the use of validated microsimulation models, before clinical- and cost-effectiveness of tomosynthesis screening versus mammography screening can be fully characterized.

Keywords: cost-effectiveness, mammography, screening, tomosynthesis

DOI:10.2214/AJR.16.16642

Received April 29, 2016; accepted after revision June 10, 2016.

This article is a commentary on "Cost-Effectiveness of Tomosynthesis in Annual Screening Mammography" by Kalra et al.

C. I. Lee was supported in part by an American Cancer Society grant (126947-MRSG-14-160-01-CPHPS). All authors were supported in part by a National Institutes of Health (National Cancer Institute) grant (P01 CA154292).

C. I. Lee and J. M. Lee receive grants from GE Healthcare that are not related to this work.

[1]Department of Radiology, University of Washington School of Medicine, 825 Eastlake Ave East, G3-200, Seattle, WA 98109-1023. Address correspondence to C. I. Lee (stophlee@uw.edu).

[2]Department of Health Services, University of Washington School of Public Health, Seattle, WA.

[3]Hutchinson Institute for Cancer Outcomes Research, Fred Hutchinson Cancer Research Center, Seattle, WA.

[4]Department of Medicine and The Dartmouth Institute for Health Policy & Clinical Practice, Norris Cotton Cancer Center, Geisel School of Medicine, Lebanon, NH.

AJR 2016; 207:1–3

0361–803X/16/2075–1

Early prospective European studies suggested that combined biennial screening with digital breast tomosynthesis and mammography can significantly decrease recall rates and increase cancer detection rates relative to biennial mammography screening alone [1]. Largely on the basis of this promise, tomosynthesis has become widely adopted in the United States in recent years, obtaining U.S. Food and Drug Administration approval in 2011 and Centers for Medicare & Medicaid Services reimbursement in 2015. From 2013 to 2015, tomosynthesis availability at imaging facilities affiliated with the National Cancer Institute–funded Breast Cancer Surveillance Consortium increased from 0% to 50% [2].

The adoption of tomosynthesis into U.S. clinical practice, however, is outpacing the evidence for its clinical effectiveness and ability to improve longer-term outcomes. Currently, no large-scale, U.S.-based prospective trials are examining the intermediate and long-term outcomes associated with tomosynthesis screening. Given its rapid diffusion into community practices, projections regarding the potential added value of tomosynthesis, namely decreases in mortality and morbidity relative to its additional costs, would help third-party payers make insurance coverage decisions and help policy makers issue screening recommendations for the general population.

In this issue of *AJR*, Kalra et al. [3] present a Markov cohort decision-analytic model comparing annual mammography screening with and without tomosynthesis for U.S. women 40 years old and older. The study team found that the incremental costs per quality-adjusted life year gained of adding tomosynthesis to annual mammography screening were well under the generally accepted threshold for cost-effectiveness (< $100,000) across all age subgroups. Their suggestion that adding tomosynthesis to annual mammography screening is a cost-effective measure held up in 63.2% of their simulations. Can we then conclude that annual combined screening should be implemented for women 40 years old and older?

Answering this question requires understanding the criteria for assessing the quality of simulation models. Models attempt to simplify features of complex real-world situations into mathematic form and to draw conclusions on the basis of simulated results. The checklist items for judging the quality of models fall into three general areas: the model structure (including the underlying assumptions used to create the model), the data used as inputs for the model (including the robustness of the studies from which input parameters are taken), and model validation [4]. In all three areas, the methods and approach described by Kalra et al. [3] do not meet criteria for a rigorous model, making their results insufficient for

confirming that annual combined screening is truly cost-effective.

In addition to these three areas that define simulation model quality, cancer screening models can further be generalized as being either an empirically based shallow model or a biologically based deep model [5]. Kalra et al. [3] present a shallow model that uses observable outcomes (e.g., improved cancer detection rates) but does not specify processes that lead to those outcomes. In contrast, a deep model simulates the natural history of cancer from tumor onset through growth and metastases and considers the mechanism of preclinical disease detection by a screening test [5].

The face validity of results from a decision-analytic model can largely be judged on the robustness of the assumptions made in the development of the model and of the data sources used to inform the model [4]. Shallow stage-shift models, such as the one used by Kalra et al. [3], make assumptions that individuals start in an initial health state with differing probabilities of changing from one health state to another. In models for breast cancer, these health states usually include differing stages of disease, from having no breast cancer to ductal carcinoma in situ, locally invasive cancer, regionally invasive cancer, and metastatic disease. Costs and outcomes associated with these different cancer stages vary widely.

The model developed by Kalra et al. [3] allows only two health states to describe the stage distribution of breast cancers, either in situ disease or invasive disease. A model constructed in this manner artificially groups all invasive cancers together in terms of costs and outcomes and does not capture actual breast cancer management practices. Presumably, potential life-saving benefits from tomosynthesis screening reside in shifts between stages of invasive cancer, and these shifts are not captured in the current model. Ideally, simulation models would also undergo validation [4], a process in which trial data not used as input parameters with known outcomes are used to verify that the model produces results consistent with the real world it represents. However, such rigorous model validation was not performed by Kalra et al. [3], in part because data for validation are not available.

In a stage-shift model, performance measures that would drive the modeling results include the sensitivity and specificity of combined mammography and tomosynthesis annual screening by women's age and whether women are undergoing incidence or prevalence screens. However, the source used for performance measures was a single retrospective study of audit data across 13 imaging facilities before and after tomosynthesis implementation at the facility level [6]. Improvements in recall and cancer detection rates across facilities between two arbitrary time periods may not translate into accurate estimates for combined screening sensitivity, specificity, or interval cancer rates by age or by prevalence versus incidence screens. These values, which would have improved model estimates of outcomes, have yet to be reported in the medical literature.

Shallow stage-shift models also require the population to be similar to the study population from which input parameters are obtained for their results to be generalizable. Because the retrospective audit study was performed in aggregate and included a convenience sample of women at 13 facilities, it is not clear that it reported on a population representative of all women in the United States. Women undergoing mammography screening in the United States experience an average screening interval somewhere between 1 and 2 years. Without longitudinal long-term patient-level data, the use of this retrospective study as a data source leaves great uncertainty about the precision of the screening performance estimates used.

Moreover, this model, which focuses primarily on the benefits of tomosynthesis, likely overstates the value of tomosynthesis screening by not accounting for screening harms. The model assumes no false-negative results under tomosynthesis screening, which implies that tomosynthesis has 100% sensitivity. Yet, tomosynthesis screening has been associated with an increase in benign biopsy rates, with implications of increased downstream resource utilization and costs [6]. This model also suggests that all additional invasive cancers detected by tomosynthesis are clinically relevant. However, a proportion of screen-detected invasive cancers may represent overdiagnosis and overtreatment that would not have led to breast cancer morbidity or mortality if left undetected. This assumption of long-term benefit from detection with tomosynthesis leads to an estimate of effectiveness in the upper end of the range of plausible effectiveness, contributing to a favorable incremental cost-effectiveness ratio.

Given both the limitations of the model's underlying structural assumptions and the nature of the data source used, we believe that no reliable conclusion regarding the cost-effectiveness of annual tomosynthesis screening can be reached. Instead, the analysis by Kalra et al. [3] highlights the need for more robust U.S. population–based data from large observational studies, clinical trials, or both comparing tomosynthesis and mammography screening.

Once reliable long-term outcomes data become available, biologically based deep models can be used to compare multiple tomosynthesis screening strategies based on women's risks and screening intervals. These deep models are more comprehensive than stage-shift models and can be used to address a larger range of policy issues, including screening intervals, lead-time bias, overdiagnosis, and changes in imaging test performance over time [5].

Large-scale observational registries, such as the Breast Cancer Surveillance Consortium, can provide reliable estimates of mammography and tomosynthesis performance in the U.S. population stratified by patient risk factors and screening intervals over time. The Cancer Intervention and Surveillance Monitoring Network offers six validated deep models for breast cancer with the ability to apply a comparative modeling approach to a range of population-based screening questions.

In summary, the assumptions made in model development and the performance data source used in the study by Kalra et al. [3] limit the validity and generalizability of its results. Although tomosynthesis screening in the United States holds promise, its cost-effectiveness remains undetermined. More representative longitudinal outcomes data and simulation modeling, including the biennial screening strategies recommended by the U. S. Preventive Services Task Force and the American Cancer Society, are needed before the cost-effectiveness of tomosynthesis screening can be established.

References

1. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 2013; 14:583–589

2. Houssami N, Miglioretti DL. Digital breast tomosynthesis: a brave new world of mammography screening. *JAMA Oncol* 2016; 2:725–727

3. Kalra VB, Wu X, Haas B, Forman HP, Philpotts LE.

Cost effectiveness of tomosynthesis in annual screening mammography. *AJR* 2016; 207: in press

4. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices—modeling studies. *Value Health* 2003; 6:9–17
5. Knudsen AB, McMahon PM, Gazelle GS. Use of modeling to evaluate the cost-effectiveness of cancer screening programs. *J Clin Oncol* 2007; 25:203–208
6. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014; 311:2499–2507