Article details: 2020-0154	
Title	The cost-effectiveness of adding tomosynthesis to mammography-based breast cancer screening: an economic analysis
Authors	Sonya Cressman PhD MBA, Colin Mar MD, Janette Sam MRT, Lisa Kan MSc, Caroline Lohrisch MD, John J. Spinelli PhD MSc
Reviewer 1	Dr. Abdullah Ababousi
Institution	
General comments (author response in bold)	Thank you for the opportunity to review this manuscript titled:" The cost-effectiveness of adding tomosynthesis to mammography-based breast cancer screening" General Comments: -Very interesting topics and very relevant at this time **Thank you for this positive feedback** E.31-One thing to consider is could we focus on the cost-effectiveness of targeting baseline screening studies only. These have the highest call-back rates and are the most problematic in terms of false-positives. You noted in your study callback rates of 19.5% vs 9%. **R.31. Thank you for this suggestion. Our deterministic analysis included one scenario for baseline exams only, however in line with comments above, we have clarified these results into a tornado diagram as suggested above.** E.32-Another thing to consider is the difference in the cost of call-backs between patients receiving DM only vs DBT+DM. With DM only, the callback entails additional mammographic imaging as well as US. With DBT+DM callbacks, patients generally only require US. **R.32. This scenario would improve the cost-effectiveness of DBT by a small magnitude, and much less than the impact of the lowest cost-estimate for DBT exams. A general description of variability in costs has been added to the methods section describing the deterministic analysis (page 5, line 383).** E. 33-Finally, multiple studies are now suggesting that the synthetic 2D mammographic images generated as part of the DBT image acquisition are comparable to standard to standard 2D DM. so that question here is will screening with DBT alone (which includes 3D tomosynthesis as well as synthetic 2D images) be even more cost-effective? **R.33. The cost-effectiveness will vary depending on the recall rates with DM. Should DBT screening costs decrease (i.e. due to widespread availability of DBT alone as older DM machines are displaced), the cost-effectiveness will improve, however if the technology does not substantially reduce recall rates or if it actually increases them, the effect will be washed out by the more impactful parameter. This is illustrated in the additional Figure 2 that was added for clarification.** E.34 Abstract: -Clear and concise -Line 47: remove ‘s’ from ‘increases’ **R.34. The suggested change has been made** E. 35. Introduction:

	-Line 87: You need a reference(s) for your first sentence **R.35. The suggested change has been made (page 3, line 150)** E.36. -" lower rates of overdiagnosed breast cancer that is not life threatening": Multiple studies have actually shown that DBT is leading to the detection of less aggressive cancers, which are sometimes subtle/slowly growing over many years and difficult to otherwise detect on standard 2D DM. For example see: Biologic Profiles of Invasive Breast Cancers Detected Only with Digital Breast Tomosynthesis, Kim et al., AJR 2017. So I would remove this point, unless you have a specific reference for it. **R.36. The suggested change has been made (line 163, page 3 was removed)** Methods: -Clear. No issues. Results: -Clear. No issues. Discussion: -Clear and concise, but please see the 'General Comments' section above for additional things to consider/discuss
Reviewer 2	Dr. Paola Giorgi Rossi
Institution	Epidemiology Unit, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Emilia-Romagna, Italy
General comments (author response in bold)	E.37. The paper presents a CEA comparing DBT+DM vs DM alone. Actually DBT+DM is not recommended anymore by most recent guidelines, since the introduction of DBT+synthetic 2D have similar accuracy and gives half radiation dose than DBT+DM. Unfortunately, evidences about specificity of DBT+Synthetic 2D are weaker than those about DBT+DM and some European study suggest that there is a lack in specificity, particularly when the previous mammograms are with real DM. Thus all the rationale of the study is less sound in the light of the new recommendations. **R.37. Thank you for this comment. We have cited negative task force recommendations for the North American context on lines 175-177, page 3 of the Introduction, and in the Interpretation section (lines 585-586, page 8). In addition, the new Figure 2 illustrates the sensitivity analysis in our study and the overall impact on cost-effectiveness. Changes have been made as described in previous revisions to emphasize this main finding.** E.38. Another weakness pf the study is the use of reimbursement fees also for the cost of DBT+DM and DM. I understand that the analysis is from the health service point of view, but fees should be made according to accurate costing analyses or, if the public health system wants to use them to make policies, on the basis of CEA. Here the authors are doing the opposite, thus creating a tautological finding. This is even more critical when the technology is rapidly evolving, so if the fee would be really based on accurate costing analyses, it should be changed for DBT+Synthetic 2D, because the time for the radiographer to acquire the images is reduced. These two points, going in opposite directions, makes all the model results less useful to support decision making. **R.38. The use of reimbursement fees to estimate DBT+DM costs is a**

	standard costing approach that makes our analysis comparable with others and with the other types of costs (i.e. costs for treatment and diagnostic work-up) in the model. In the real world, DBT+DM costs will vary—for example, if screening programs are able to maximize economies of scale, if new technology competes to displace DBT or can improve data management, or if the comparative DM equipment depreciate over time and replacement costs are high. Of particular concern for Canada is the economies of scale aspect and the ability to deliver health services equitably to remote and rural areas with mobile screening vans. We believe this is addressed by the wide range of costs we tested in the deterministic sensitivity analysis, with a lowest estimate based on a published bottom-up costing method for DBT in the EU and the highest estimate based on published U.S. fees, which are typically 3-10X higher than Canadian fees for other radiologic services. E.39. Specific comments. Methods I think the model should be better explained in its mechanisms: how changes in parameters changes overdiagnosis? And how impact on mortality? **R.39. We have revised the suggested section of the methods to improve the clarity and describe the sensitivity analysis (page 5, lines 378-389).** E.40. I understand that the model does not admit regression of lesion, so the overdiagnosis is saturated after some screening rounds, how can be generated overdiagnosis by DBT? **R.40. We have added text to describe how overdiagnosis was modelled in the paragraph of the methods on Page 5 as cited above in R.3. We have also removed a sentence suggesting additional overdiagnosis to the revisions suggested by Reviewer #1 (E.36)** E.41. Is it generated only in the screening rounds close to the end of screening period? The model does not assume the presence of a pool of indolent cancers for which the DBT has higher sensitivity than DM. How all these assumptions relate with the detection and/or progression of DCIS? All these assumptions should be made explicit in the Paragraph about the model description. **R.41. We have added simplified health state definitions to the caption of Figure 1 and reference to the Supplement in the Methods section (Page 4, lines 286-287). Due to the limit on word counts and the large amount of information that goes into economic models, we felt that the required health state transitions and diagnostic grouping information is be best explained in detail in the Supplementary Materials.** E.42. I did not find the definition of high and low risk, sorry if I missed it. **R.42. The suggested change has been made to the caption of Figure 1.** E.43. A better description of the screening protocol or practice would help the reader to understand the real interventions you are comparing: age interval 40-74, but screening for 23 years… **R.43. The suggested clarification has been made to the first paragraph of the**

	Methods section. E.44. Please try to compare fees with costs from data bout radiographer time for acquisition, radiologist time for reading, technology costs for updating machines and particularly electronic image storage, and all the other costs from that can differ based on a literature search. This would make the sensitivity analysis able to account for the introduction of synthetic 2d at least for the cost parameters. **R.44. We address this aspect of the modelling with a deterministic sensitivity analysis shown in the new Figure 2.** E. 45 Discussion It is important to discuss the uncertainty about how the increase in sensitivity will impact on mortality reduction and on overdiagnosis. Initial data on interval cancers show small, if any, reduction, suggesting that the impact on mortality and prognosis could be modest or null. **R.45. The suggested change has been made to the Discussion and throughout the revised manuscript as explained in Revision #1** E.46. The improvement in specificity seems to be strongly context dependent and it should be mentioned that in European studies the recall rate was similar with DBT+DM and with DM. The recent systematic review made for the European Guidelines (ECIBC) suggests that when the recall rate is low with DM it is difficult that DBT may reduce it, on the contrary when it is high (as in most US studies and in your British Columbia cohort) DBT will probably be more specific than DM. **R.46. We agree and have made the suggested change throughout the revised manuscript.** **In addition, minor changes are noted throughout the markup versions were made to the manuscript to improve the readability and remain within the suggested word limits.**