Peer Review File

Article information: https://dx.doi.org/10.21037/gs-21-599

Reviewer A

Comment 1: The authors have analyzed the possible relationship between Histopathological Growth Distribution of Ductal Carcinoma in Situ and the development of distant metastatic disease. It is apparently a good and original idea. However, there is a significant limitation in the study. A total of 226 patients were included in the paper, but half of the data lacked Histopathological Growth Distribution. In addition, the majority of diffuse lesions were missing size data. The author should add more data to prove the thesis.

Reply 1: We thank Reviewer A for their careful read of our manuscript and for their feedback. We agree with Reviewer A that the large number of missing size data among patients with diffuse lesions is a significant limitation to our study. However, this is actually part of the reason for our study—different pathologists at our institution, as well as several other institutions we have collaborated with, document size for these diffuse lesions differently, and we hope with this study to suggest the need for documentation of both the size of the greatest focus of disease, as well as the overall extent of the lesion, based on the pathology analysis. We suspect that among diffuse type DCIS lesions, the overall extent of disease, rather than the largest focus of disease as would appear in national databases such as SEER data, is the most important prognostic factor. Unfortunately, since only representative slides and blocks are stored long term by the pathology department at our institution, even if slides and blocks were available for all of our patients we would be unable to estimate the overall extent size of the diffuse lesions from this. As a result, we are limited to the data available from the pathologists' reports so there is unfortunately no method to obtain additional information on the extent of disease for the patients studied.

The alternative to obtaining more data would be to include more patients in our study. Our IRB allowed us to look at patients from January 1, 2000 to March 21, 2019. However given that this roughly 19 year period results in our identification of 259 patients meeting criteria, it is unlikely that writing up a new IRB such that we could include patients form the last two and a half years would markedly change the study size. Furthermore, we would not gain any more information on the outcomes of the patients included in our study since we chart reviewed all of them, so we wouldn't have additional patients suffering breast cancer metastasis. Also of importance, if we were to include these patients, they would have a maximum of two and a half years follow up, so it is unlikely that this cohort would contribute to the analysis of progression to metastatic disease as the median time among patients suffering this outcome was 6.6 years and mean time 4.3 years.

The other alternative to increase the sample size would be to conduct a multi-institutional study. We have collaborated with other institutions on DCIS projects, however to conduct this study we would need an independent, multi-institutional IRB, as well as IRBs from all other participating institutions, and

funding for associated team members to perform extensive chart review. While this has the potential to increase both our overall sample size and the number of patients with histopathological diffuse type distribution disease, pursuing this route would require at least another year of work, and we do not have funding for teams at other institutions to complete this review. It is our hope that the publication of this manuscript will generate interest in this topic among both pathologists and surgeons so that we might attract partners to work with us on the next steps of our investigation into diffuse type disease which would involve prospective large format histology of large DCIS lesions to clarify the invasive/microinvasive content in a more definitive manner and allow for the potential to perform single cell genetic analyses on different foci as well as transcriptomics and to assess for clonal relationships among foci, as we suspect that diffuse type disease is related to genetic predisposition to breast neoplasms and that as such the foci are less likely to be clonally related than those in multifocal or unifocal lesions.

Changes in the text: We added the following to the discussion portion of the text on page 8: "Another limitation is the large number of missing largest contiguous size data for the diffuse type group. We were limited to what was available in the pathology reports since only representative specimens from the original surgical specimen are stored long term, preventing us from estimating both largest contiguous size and extent of disease."

Reviewer B

Comment 1: Interesting paper well conducted.
should note that some registries consider node positive dcis as not dcis and reclassifies these as stage 2. what are the implications of this

Reply 1: We thank Reviewer B for their careful read of our manuscript and thoughtful suggestions. This is a great observation regarding microinvasive disease. And they are correct, our registry does document such patients as stage 2. Since our registry specifies microinvasive disease in the pathologic and clinical T stages, we were able to identify these patients independent of their overall stage and so patients with this type of disease were not excluded from our study.

From the standpoint of clinical implications, the question would be whether the practitioners treating a given patient changed their approach to systemic therapy on this basis. At our institution, none of the patients, including those with microscopic or macroscopic foci of lymph node involvement, identified for this study received chemotherapy, independent of the molecular subtype identified in the lesion. While none of the three patients in our registry identified as having lymph node metastasis or micrometastasis went on to develop metastatic breast cancer, it should be noted that two of the three of them received adjuvant endocrine therapy, as opposed to none of the five patients identified as developing metastatic breast cancer, so it is unclear whether this is related to the increased use of adjuvant endocrine therapy in this group (notably three of five patients in the metastatic group had hormone receptor negative

disease and would not meet criteria for endocrine therapy) or to a prognostic difference in lymph node metastasis in DCIS vs. diffuse type DCIS.
Changes in the text: We added the following to page 7 in the discussion: "While none of the patients with lymph node metastases in our study were treated with adjuvant chemotherapy, it is worth noting that such clinical scenarios can be categorized as Stage IIA lesions. Such patients could then potentially be considered for adjuvant chemotherapy depending on the subtype, though given the non-overlap in our cohort of development of subsequent metastatic disease with identification of sentinel lymph node metastasis, our results do not provide evidence to support this."

Comment 2: need a proper survival analysis. need to include micro-invasion as co-variate
were the metastases limited to those with micro-invasion? this is important
Reply 2: We agree with Reviewer B that a survival analysis incorporating appropriate covariates would be ideal. Unfortunately, due to the small number of events of interest and small cohort size, this is not possible—we attempted several proportional hazards survival models, but even the univariate models by subtype were too small to converge.
Changes in the text: We added to page 7 "Another limitation is that due to the small sample size and small number of events of interest, we were unable to construct a proportional hazards model for development of subsequent metastatic disease as such a model did not converge."

Comment 3: expand this to include snld, growth pattern and presence of micro-metastases lines 200-202.The main outcome should be metastatic disease not positive nodes
Reply 3: As mentioned in reply 2, due to the small sample and event sizes, we were unable to perform a proportional hazards survival analysis, and were limited to Kaplan-Meier estimators. As a result, we could not adjust for sentinel lymph node involvement or presence of micro-metastases. However, as is noted on page 5, none of the patients with sentinel lymph node involvement went on to develop metastatic disease.
Changes in the text: We added to page 7 "Another limitation is that due to the small sample size and small number of events of interest, we were unable to construct a proportional hazards model for development of subsequent metastatic disease as such a model did not converge."
Comment 4: what proportion of those who developed metastatic disease had positive lymph nodes five of how many line 186 is misleading and is inconsistent with line 191
Reply 4: None of the patients who developed metastatic disease had positive lymph nodes. In the introduction section we define SLNI/DMD as "sentinel lymph node involvement or the development of subsequent metastatic disease (SLNI/DMD)", so this is a composite outcome for either event. This was for matters of practicality related to the small event number—we agree completely with Reviewer B that the

preferable analysis would be a multivariate proportional hazards model incorporating the relevant covariates, this is just unfortunately not analytically possible with our small sample size and small event number.
Changes in the text: We have edited page 5 to reiterate the definition of SLNI/DMD: "Either sentinel lymph node involvement or presence of distant metastatic disease (SLNI/DMD) was identified in 9 (4.0%) of the 226 patients (Table 1), including 5 (2.2%) patients who developed distant metastatic disease and 4 (1.8%) who were found to have isolated tumor cells or micrometastases on sentinel lymph node biopsy during one or more of their DCIS procedures."

Comment 5: study sample small, any way to enlarge it of get more missing data?
Reply 5: This is an excellent question but there is unfortunately not a way for us to obtain more of the missing data. As discussed in response to Reviewer A (edited slightly from above response):
The large number of missing size data among patients with diffuse lesions is a significant limitation to our study. However, this is actually part of the reason for our study—different pathologists at our institution, as well as several other institutions we have collaborated with, document size for these diffuse lesions differently, and we hope with this study to suggest the need for documentation of both the size of the greatest focus of disease, as well as the overall extent of the lesion, based on the pathology analysis. We suspect that among diffuse type DCIS lesions, the overall extent of disease, rather than the largest focus of disease as would appear in national databases such as SEER data, is the most important prognostic factor. Unfortunately, since only representative slides and blocks are stored long term by the pathology department at our institution, even if slides and blocks were available for all of our patients we would be unable to estimate the overall extent size of the diffuse lesions from this. As a result, we are limited to the data available from the pathologists' reports so there is unfortunately no method to obtain additional information on the extent of disease for the patients studied.

The alternative to obtaining more data would be to include more patients in our study. Our IRB allowed us to look at patients from January 1, 2000 to March 21, 2019. However given that this roughly 19 year period results in our identification of 259 patients meeting criteria, it is unlikely that writing up a new IRB such that we could include patients form the last two and a half years would markedly change the study size. Furthermore, we would not gain any more information on the outcomes of the patients included in our study since we chart reviewed all of them, so we wouldn't have additional patients suffering breast cancer metastasis. Also of importance, if we were to include these patients, they would have a maximum of two and a half years follow up, so it is unlikely that this cohort would contribute to the analysis of progression to metastatic disease as the median time among patients suffering this outcome was 6.6 years and mean time 4.3 years.

The other alternative to increase the sample size would be to conduct a multi-institutional study. We have collaborated with other institutions on DCIS projects, however to conduct this study we would need an independent,

multi-institutional IRB, as well as IRBs from all other participating institutions, and funding for associated team members to perform extensive chart review. While this has the potential to increase both our overall sample size and the number of patients with histopathological diffuse type distribution disease, pursuing this route would require at least another year of work, and we do not have funding for teams at other institutions to complete this review. It is our hope that the publication of this manuscript will generate interest in this topic among both pathologists and surgeons so that we might attract partners to work with us on the next steps of our investigation into diffuse type disease which would involve prospective large format histology of large DCIS lesions to clarify the invasive/microinvasive content in a more definitive manner and allow for the potential to perform single cell genetic analyses on different foci as well as transcriptomics and to assess for clonal relationships among foci, as we suspect that diffuse type disease is related to genetic predisposition to breast neoplasms and that as such the foci are less likely to be clonally related than those in multifocal or unifocal lesions.

Changes in the text: We added the following to the discussion portion of the text on page 8: "Another limitation is the large number of missing largest contiguous size data for the diffuse type group. We were limited to what was available in the pathology reports since only representative specimens from the original surgical specimen are stored long term, preventing us from estimating both largest contiguous size and extent of disease."

Comment 6: should comment that mastectomy doesn't reduce mortality compared to lumpectomy in dcis but radiation does
see Narod, Giannakeas papers in JAMA Oncology and JAMA network open

Reply 6: This is an interesting point from Reviewer B. It is true that Narod and Giannakeas found reduced mortality with adjuvant radiation but not mastectomy. Importantly, however, in the JAMA Oncology paper referenced patients undergoing bilateral mastectomy were excluded, and in the JAMA Network paper they did not explicitly state whether only unilateral cases were included as in their prior study or if both were included. Due to the method by which SEER coders are instructed to recode initially in situ lesions to invasive if there is a subsequent metastatic event in the absence of an intervening invasive event, as roughly half the patients Narod and Giannakeas identified were, this will naturally lead to an underestimation of mortality for patients who undergo mastectomy, as we have previously published on ("Surveillance, Epidemiology, and End Results program underestimates breast cancer-specific mortality after ductal carcinoma in situ diagnosis", Breast Cancer Research and Treatment.) The other problem is that both Narod and Giannakeas used tumor size as coded by SEER, which for their registrars is defined as the largest contiguous focus of disease, so as a result any patients with diffuse disease as in our study would likely be coded as a small tumor size, as the diffuse type pattern involves numerous small foci spread out over a large lesion. Therefore even with propensity score matching as they performed in the JAMA Network paper would not properly account for patients with diffuse type disease and would mistakenly match them

size-wise with a patient with a small, likely unifocal lesion. Therefore we will add a comment about Narod and Giannakeas's studies, but also a warning about these results as we suspect that mastectomy may be superior for these patients with diffuse type disease, as we suspect a genetic predisposition among these patients.

Changes in the text:

"Narod and Giannakeas have previously demonstrated using SEER data that patients undergoing unilateral mastectomy did not have a survival benefit over patients undergoing lumpectomy, whereas patients undergoing lumpectomy with adjuvant radiation therapy did. Problematically, since SEER registrars are instructed to code largest contiguous focus of disease rather than overall extent of disease, this would lead to patients with diffuse DCIS to being matched on the basis of their largest focus of disease, which could be quite small, and would therefore likely mismatch these patients. Furthermore, SEER registrars are instructed to recode patients who suffer breast cancer mortality after DCIS without an intervening invasive lesion as having initially had an invasive lesion, which is inherently problematic in terms of using SEER to estimate breast cancer mortality after a diagnosis of pure DCIS. Furthermore, we showed in a recent analysis including patients undergoing both unilateral and bilateral mastectomy that while in a univariate model mastectomy did not reduce breast cancer mortality, in the multivariate model there was a significant reduction, though this conclusion is still limited by the previously mentioned shortcomings of SEER data."

Comment 7: This is a very important observation and we need more work in this area.

is there literature on growth distribution and local recurrence? if so please quote

Reply 7: The only evidence we are aware of for this was produced by Tibor Tot, in a small case series, which didn't mention statistical differences but did demonstrate what appear to be marked differences in local recurrence rates. We will reference this.

Changes in the text: We added the following to page 7: "Tot demonstrated that among patients with neogenetic diffuse DCIS there is a higher local recurrence rate (27%) than among diffuse, non-neogenetic (14%), multifocal (14%) and unifocal (2%), where neogenetic refers to evidence of abnormalities in arborization of the lobe [22]."