Imputing ER Status under a Missing Not At Random (MNAR) Assumption

Dr. Rebecca Andridge
andridge.1@osu.edu

Updating ER Imputation

**Previous Analysis**
- SEER-13
- Female breast cancer patients
- Diagnosed 1992-2007
- N = 401,741
- Multiple imputation
  - Assuming MAR (missing at random)
  - Sequential regression method
  - 12 additional variables for imputation model

**Updated Analysis**
- SEER-13
- Female breast cancer patients
- Diagnosed 1992-2012
- N = 547,886
- Multiple imputation
  - Allow for MNAR (missing NOT at random)
  - Sequential regression method
  - 13 additional variables for imputation model
Missingness of ER Status

- Proportion missing ER status vs. year from 1992 to 2012 for White and Black populations.
- New data: not much missing ER.

Changes in Variables Used

**New Analysis:**

<table>
<thead>
<tr>
<th>Demographic Variables</th>
<th>Clinical Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis</td>
<td>Node positive status</td>
</tr>
<tr>
<td>Year of diagnosis</td>
<td>Metastasis at diagnosis</td>
</tr>
<tr>
<td>Registry</td>
<td>PR.Status</td>
</tr>
<tr>
<td>Race</td>
<td>Histology</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Tumor Grade</td>
</tr>
<tr>
<td>County level poverty</td>
<td>Tumor Size</td>
</tr>
</tbody>
</table>

- Why drop PR status?
  - PR status almost always missing when ER status missing
  - 99.5% of cases with missing ER also had missing PR!
MNAR Imputation for ER Status

• Why possibly missing NOT at random?

• If ER+ and ER- cases have different clinical presentation, might be more/less likely to have ER testing done
  • ER testing not part of routine clinical practice guidelines before 1996
  • Not all patients referred for ER testing
    • Maybe only those undergoing chemo
    • ER+ tumors more likely caught by screening

• Thus missing ER might be related to ER status, i.e., ER- more likely to be missing

MNAR Imputation

• Standard (MAR) imputation assumes:
  • Probability ER missing = explained completely by other variables (e.g., age, race, year…)

• MNAR imputation assumes:
  • Probability ER missing = at least partially explained by ER status itself

• We use a sensitivity parameter, \( \lambda \) (“lambda”), to capture how much we assume missing ER depends on ER
  • \( \lambda=0 \) \text{→ missing doesn’t depend on ER} \text{→ MAR}
  • \( \lambda=\infty \) \text{→ missing totally depends on ER} \text{→ MNAR}
  • \( \lambda=1 \) \text{→ missing partially depends on ER} \text{→ MNAR}
Percent ER-, Before/After Imputing

Proportion of cases that are ER- goes UP if missingness depends on ER status

Incidence, Before/After Imputing
Comparing the Imputation Results

• Complete Case vs. Imputed Data
  • Complete case underestimates incidence rates, regardless of MAR or MNAR imputation
  • Very different APCs and joinpoints in some cases
  • Underestimation less severe post-2003 (less missing!)

• MAR vs. MNAR Imputations
  • After 2003, nearly indistinguishable
  • Under MNAR, more missing cases are imputed as ER-
    • MNAR $\rightarrow$ higher rates for ER- tumors compared to MAR
    • MNAR $\rightarrow$ lower rates for ER+ tumors compared to MAR
  • But, differences not that big:
    • E.g., Black females in 1995:
      MAR = 81.0 cases per 100,000 women (SE = 3.0)
      MNAR = 87.4 cases per 100,000 women (SE = 3.4)

Conclusions

• Relatively small differences for MAR vs. MNAR
  • Joinpoint trend lines very close
  • APCs similar

• Big difference is Complete Case vs. Imputed!

• Can’t know if MAR or MNAR is “right”
  • But, this sensitivity analysis shows that using MAR assumption wouldn’t be too wrong if MNAR is really true
  • Anyway, much better than Complete Case!
Future Work

• Use same methodology to impute HER2 status
• Develop R package to allow others to use this MNAR imputation tool
  • Currently, code works with MICE package, but not in “production” form

• Reference:
  • Andridge RR, Noone AM and Howlader N. *Imputing estrogen receptor (ER) status in a population-based cancer registry: A sensitivity analysis*. Statistics in Medicine, Under Revision