To Bootstrap, or Not to Bootstrap Before Multiple Imputation: That Is the Question.

Rose Gagnon[1,2,3], Simon LaRue[2], Kadija Perreault[1,3], Luc J. Hébert[1,3], Jason R. Guertin[1,2]

Contact: rose.gagnon.1@ulaval.ca

Affiliations

1. Université Laval
2. CHU de Québec Université Laval
3. Cirris

MSR94

Context

Multiple imputation (MI) is an increasingly popular method for dealing with **missing data**[1]

In cost-effectiveness analyses, bootstrap and MI are required to **derive cost-effectiveness planes**

However, the order in which bootstrap and MI should be done <u>remains unclear</u>

First Bootstrap, then MI could yield more[1,2]:

- Valid variance estimates
- Robust inference

Especially under **non-ideal conditions**

First MI, then bootstrap is[1,3,4]:

- Computationnally **efficient**
- Supported by **standard software**

Which approach is the most robust, efficient and statistically valid?

Objective

Determine whether the **order** in which **bootstrap and MI** are performed **influences cost-effectiveness results** obtained.

Results

<u>Complete case analysis</u>

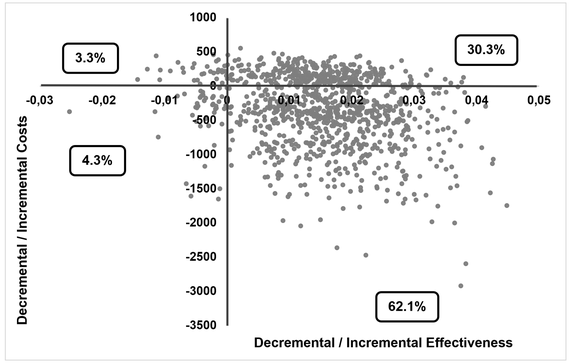

Figure 1. Cost-effectiveness plane – Canadian Public Payer perspective

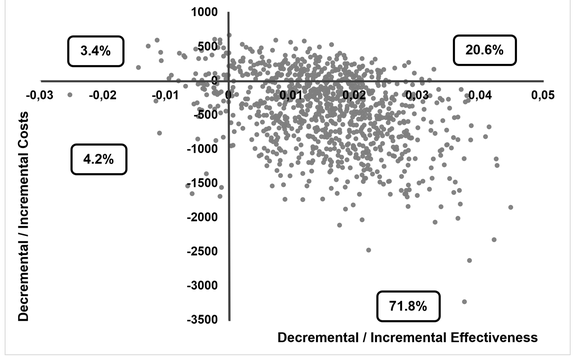

Figure 2. Cost-effectiveness plane – Canadian Societal perspective

<u>Bootstrap then multiple imputation (MI)</u>

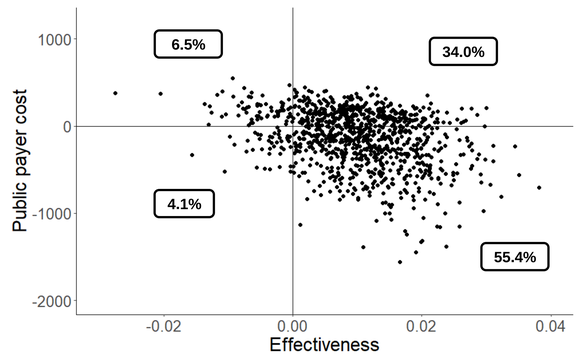

Figure 3. Cost-effectiveness plane – Canadian Public Payer perspective

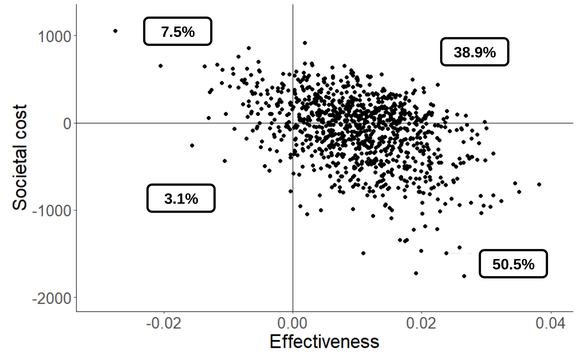

Figure 4. Cost-effectiveness plane – Canadian Societal perspective

<u>Multiple imputation (MI) then bootstrap</u>

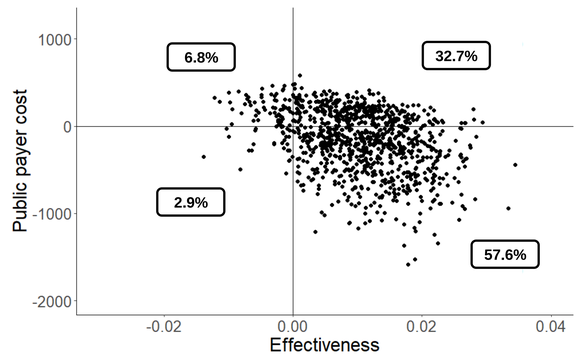

Figure 5. Cost-effectiveness plane – Canadian Public Payer perspective

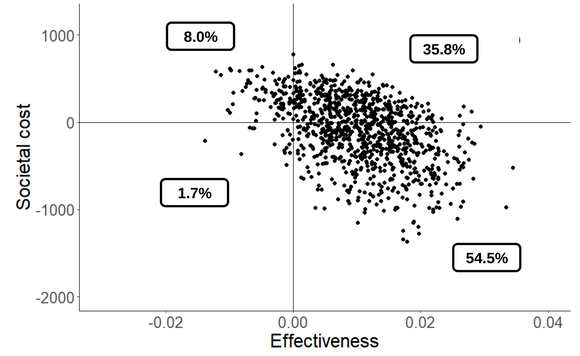

Figure 6. Cost-effectiveness plane – Canadian Societal perspective

Methods

- **Secondary analyses** using **cost** and **effectiveness data** obtained during a pragmatic **randomized clinical trial**
- ED of the CHU de Québec – Université Laval (Quebec, Canada, #NCT04009369)
- People aged **18 to 80** presenting to the ED with a **minor musculoskeletal disorder** (n=78, data missing at random)
- Analysis performed on R software according to **two scenarios**:
  - Base sample was **bootstrapped** (1,000 samples, n=78 per sample) **THEN imputed** using the MICE package (number of imputations according to % of missing data, predictive mean matching method for continuous variables)
  - Base sample was **imputed THEN bootstrapped** using the same methods
- **Cost-effectiveness planes** obtained using each scenario were **compared** to the **complete case analysis** (reference scenario)

Discussion + Conclusion

Sequence in which **MI** and **bootstrap** were used **did not have a significant effect** on results

... However, computation times were **very different** (minutes vs hours)

Depending on **statistical software**, use of the different procedures may be **less straightforward**

Further studies are needed to ascertain these conclusions:

- Different **missing data patterns**
- **Larger** sample size

<u>Limit</u>

High variability in cost and effectiveness measures

References

1. Bartlett and Hughes. *Bootstrap inference for multiple imputation under congeniality and misspectification.* 2020. https://doi.org/10.1177/0962280220932189
2. Khan and Mihailidis. *Bootstrapping and multiple imputation ensemble approaches for classification problems.* 2019. https://doi.org/10.3233/JIFS-182656
3. Wu and Jia. *A New Procedure to Test Mediation With Missing Data Through Nonparametric Bootstrapping and Multiple Imputation.* 2013. https://doi.org/10.1080/00273171.2013.816235
4. Brand, van Buuren, le Cessie and van den Hout. *Combining multiple imputation and bootstrap in the analysis of cost-effectiveness trial data.* 2018. https://doi.org/10.1002/sim.7956

Funding

As a graduate student, Rose Gagnon received financial support from the following organizations: CIHR, FRQ-S, Unité de soutien SSA Québec, Université Laval and Cirris.