

Andrea Parodi

Head of Biostatistics, EVP
Valos S.r.l.



Giacomo Siri, PhD

Principal Biostatistician
Dompé farmaceutici S.p.A.





Comparison between conditional mean and multiple imputation for Jump to Reference approach

J2R approach

Goal of the Method

- To handle **missing data** in clinical trials, especially when patients **discontinue treatment**. The **Jump to Reference (J2R)** method assumes that after discontinuation, the patient behaves like someone in the **control group**.

Conditional mean imputation

Received: 25 September 2021

Revised: 17 March 2022

Accepted: 26 April 2022

DOI: 10.1002/pst.2234

MAIN PAPER

WILEY

Standard and reference-based conditional mean imputation

Marcel Wolbers¹  | Alessandro Noci¹ | Paul Delmar¹ | Craig Gower-Page¹ |
Sean Yiu² | Jonathan W. Bartlett³

Conditional mean imputation

Step 1: Fit the Imputation Model

- Use a **Mixed Model for Repeated Measures (MMRM)** on observed data.
 - Include:
 - Treatment group
 - Visit (categorical)
 - Treatment-by-visit interactions
 - Baseline covariates
 - Estimate:
 - Regression coefficients (β)
 - Covariance matrix (Σ)
- ```
ods table SolutionF=SF;
proc mixed data=dat;
class TRT01P usubjid avisitn;
model chg=TRT01P TRT01P*avisitn avisitn
base / OUTF=outp s;
repeated avisitn/subject=usubjid type=un;
lsmeans TRT01P*avisitn/diff cl;
run;
```
- Trick: ensure the control trtgroup and the AVISIT of interest are the reference categories

# Conditional mean imputation

## Step 2. Define the Imputation Strategy

- For patients who discontinued treatment:
  - Use their expected outcomes under the treatment group **up to the point of discontinuation**.
  - From that point onward, switch to the expected outcomes as if they were in the **control group**.
- This creates a hybrid trajectory that reflects the J2R assumption: **benefit stops immediately after discontinuation**.

# Conditional mean imputation

## Step 2.

```
data SF;
set SF;
where Effect='TRT01P' and TRT01P='Active_treatment';
keep Estimate;
run;
```

```
data outpm;
set outp;
if _N_=1 then set SF;
if resid=. and TRT01P= 'Active_treatment' then do;
 chg=Pred-Estimate;
end;
if resid=. and TRT01P^= 'Active_treatment' then do;
 chg=Pred;
end;
run;
```

# Conditional mean imputation

## Step 3: Impute Missing Values (Conditional Mean Imputation)

- In this step, missing values are filled in using a **deterministic approach**—but it's important to clarify that this is **not a simplistic single imputation**. Instead, it is **mathematically equivalent** to performing **maximum likelihood-based multiple imputation** with an **infinite number of random imputations**.
- Traditional multiple imputation methods generate many versions of the dataset with randomly imputed values and then average the results. The conditional mean imputation method skips the random generation and directly calculates the expected values based on the model.
- This leads to **replicable results** (not linked to a seed) and is **computationally efficient**, while still capturing the same statistical properties as the more complex multiple imputation process.
- Although only one imputed dataset is created, it reflects what you would get if you ran thousands of random imputations and averaged them

# Conditional mean imputation

- **Step 4. Analyze the Completed Dataset**
- Once the missing values have been imputed using the conditional mean approach, the timepoint of interest in the completed dataset is analyzed using a **standard ANCOVA model**.
- It compares the treatment and control groups, adjusting for baseline characteristics.
- This step provides the **estimated treatment effect**—that is, how much better (or worse) the treatment group performed compared to the control group, after accounting for missing data and baseline differences.

# Conditional mean imputation

- Step 4.

```
proc mixed data=outpm;
 where avisit='v5';
 class trtgroup;
 model chg=baseline trtgroup/s;
 lsmeans group/diffs cl;
run;
```

# Conditional mean imputation

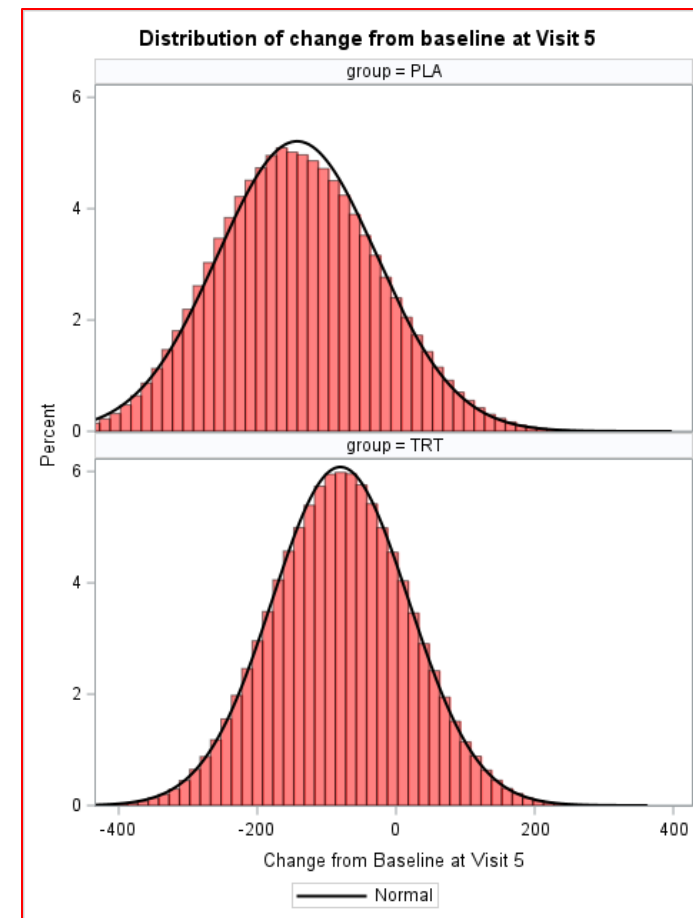
- **Step 5. Estimate Uncertainty**
- To calculate confidence intervals and p-values, use resampling methods:
  - **Jackknife**: systematically leave out one subject at a time and re-run the analysis.
  - **Bootstrap**: randomly resample the data many times and re-run the analysis.

# Creation of populations

| Timepoint | Mean in Treatment | Mean in Placebo |
|-----------|-------------------|-----------------|
| Visit 2   | -20               | -30             |
| Visit 3   | -40               | -65             |
| Visit 4   | -60               | -105            |
| Visit 5   | -80               | -145            |

% of dead subjects at v5  
5% in placebo  
1% in treated

Mean difference at v5  
 $145 - 80$   
 $=$   
 $65$



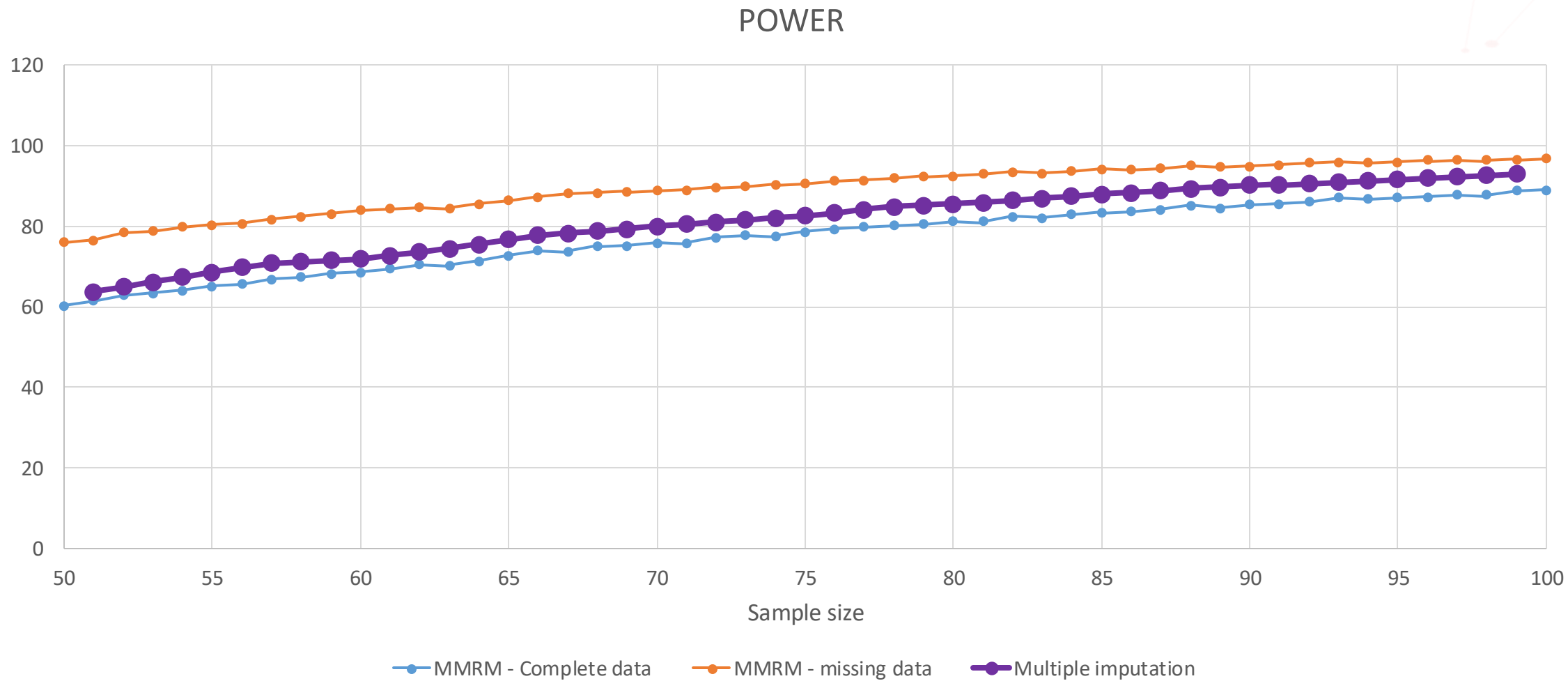
# Creation of populations with J2R

20% missingness = equal distribution between visits

|     |    | Int v1-v2 | Int v2-v3 | Int v3-v4 | Int v4-v5 | No int |
|-----|----|-----------|-----------|-----------|-----------|--------|
| TRT | V2 | -30       | -20       | -20       | -20       | -20    |
| TRT | V3 | -65       | -65       | -40       | -40       | -40    |
| TRT | V4 | -105      | -105      | -105      | -60       | -60    |
| TRT | V5 | -145      | -145      | -145      | -145      | -80    |

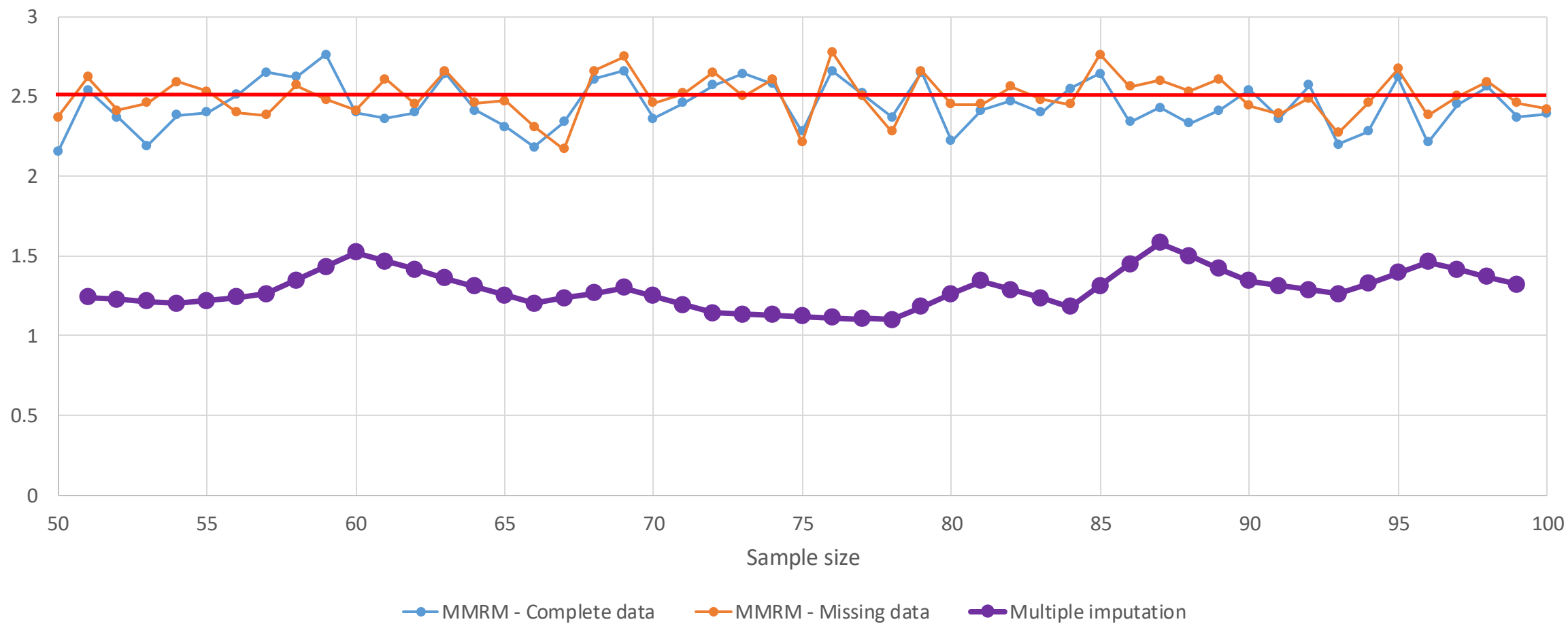
Mean difference at v5  
 $145 - 92$   
 $=$   
 $52$

# Power

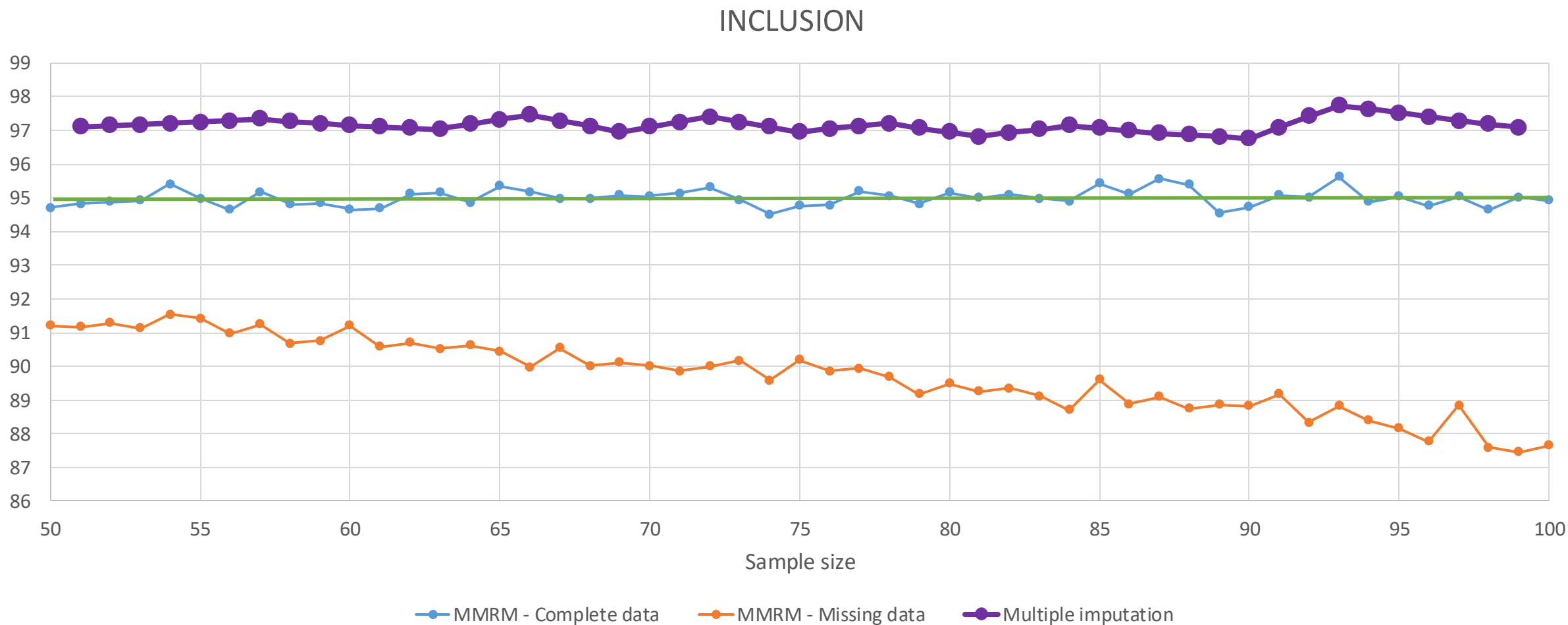


# Type 1 error

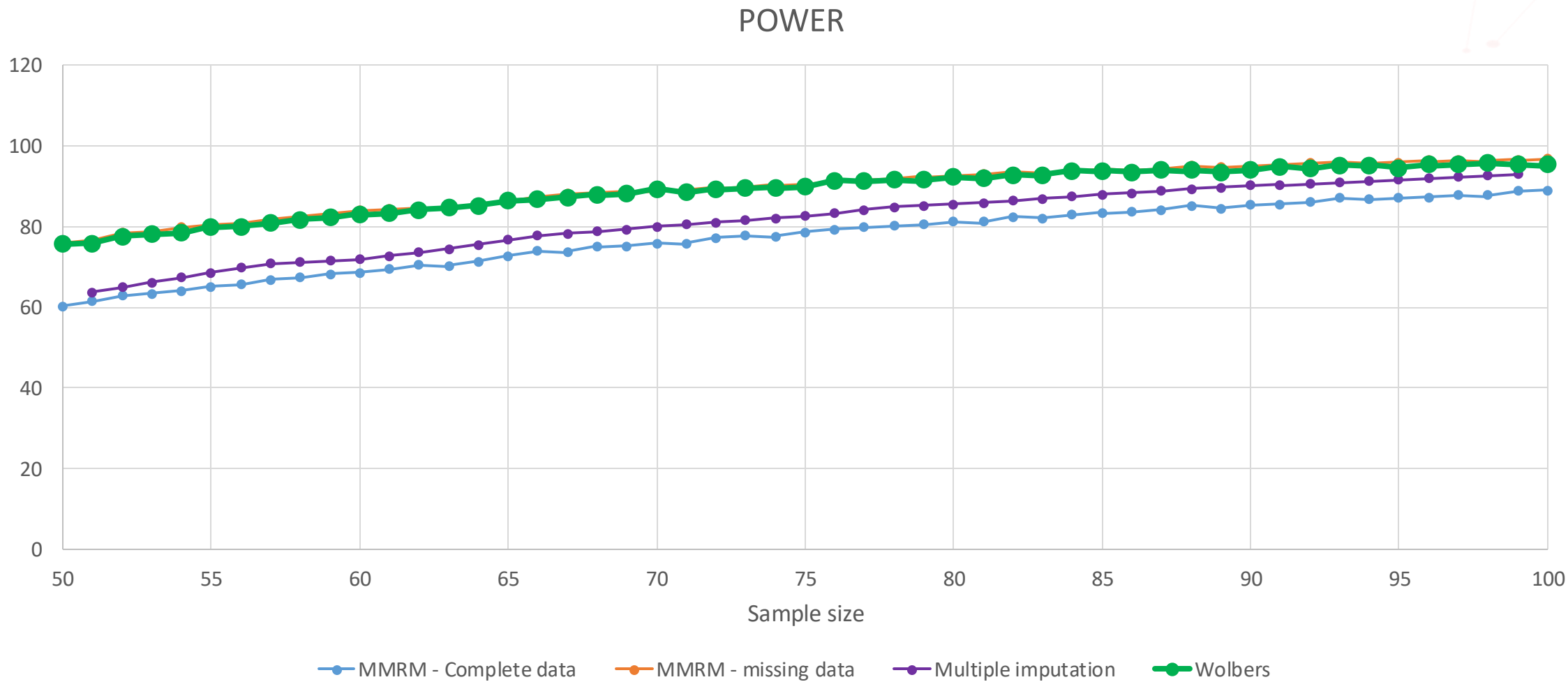
TYPE 1 ERROR



# Inclusion of the true value in the 95% CI

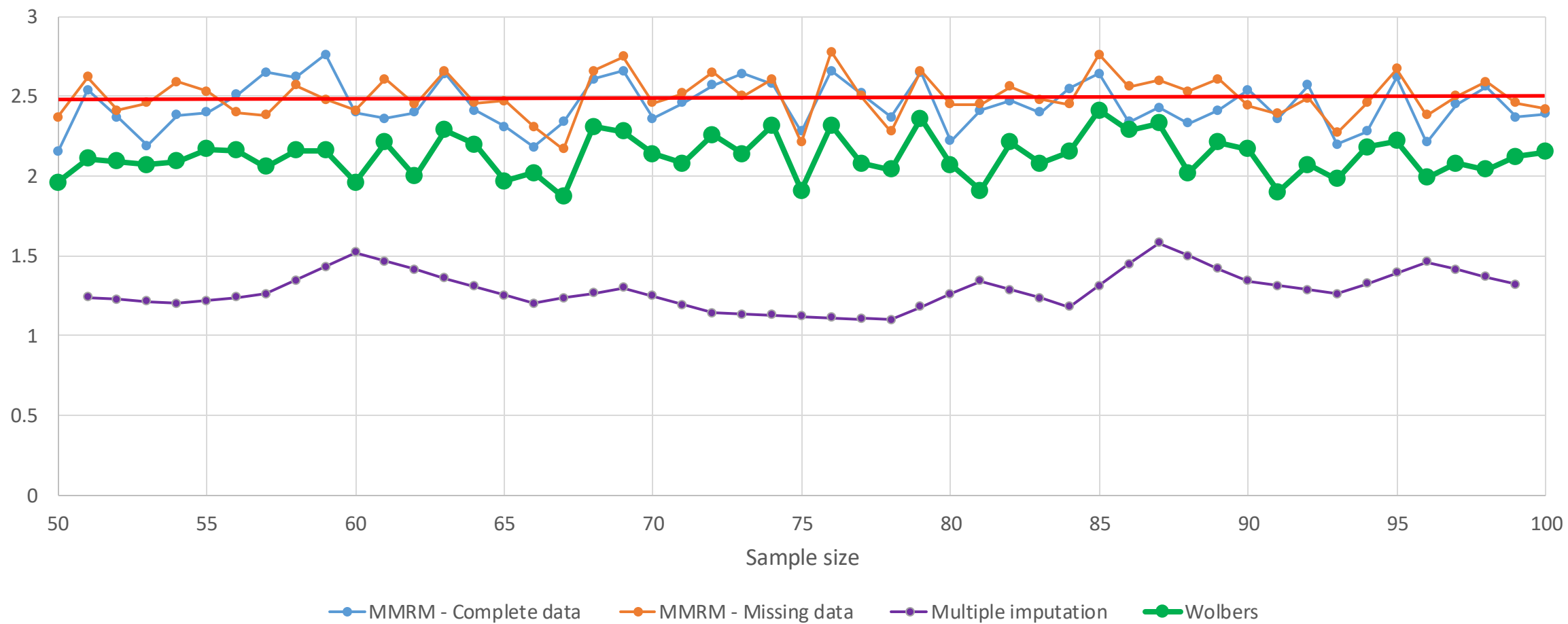


# Power

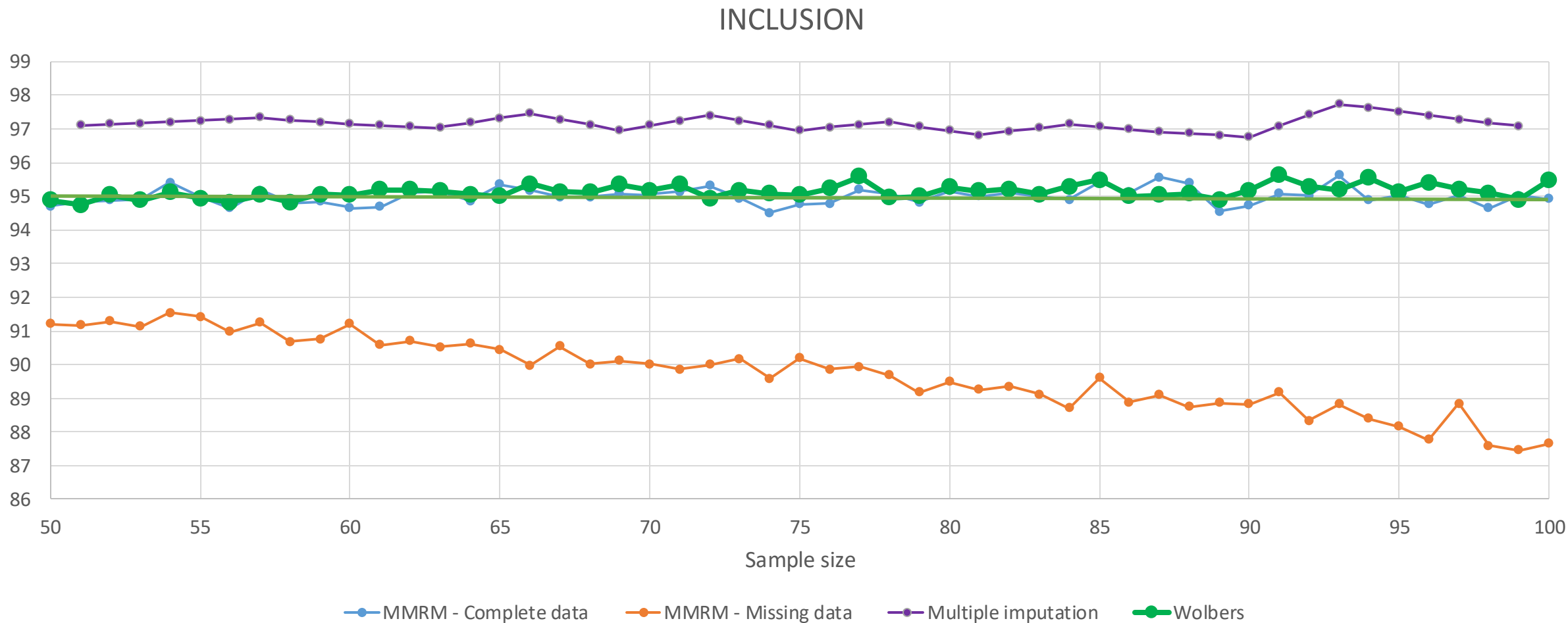


# Type 1 error

TYPE 1 ERROR



# Inclusion of the true value in the 95% CI



## Conclusions

- Multiple imputation and conditional mean imputation are both reliable
- Type 1 error is preserved
- Conditional mean imputation is faster and does not rely on any seed

## Limitations

- The simulated distributions are symmetric
- Limited number of iterations
- Not all points from 50 to 100 being simulated for MI

# Thanks for your attention

For any question please contact me at:

[andrea.parodi@valos.it](mailto:andrea.parodi@valos.it)