IMPC statistical pipeline check list

# Step1. Reading and preparing the Parquet files

* Run the Parquet to Rdata job creator
* Run Parq2Rdata function
* Run MergeRdata job creator
* Run MergeRdata function
* Clean up the unnecessary Rdata files and directories
* Compress the log files

# Step2. Generating the input data

* Create the jobs for the InputDataGeneration process (script)
* Set the memory/CPU etc and run the InputDataGeneration script

# Step3. Create the jobs

* Collect all single jobs and merge them to create the Alljobs.bch (make sure memory and CPU are correctly set)
* Compress the single files and remove the redundancies

# Step4. Preparing the statistical engine

* Update the IMPReSS related file and replace in the R package source +exception list
  + Can be run once on the cluster
* Update all packages from GitHub
  + Note that if do not want to run the update from the cluster then must be first updating the Impress files and replace in the package source (committed) and then run the update script
* Update optional and skip list
* Check the function.R script for the analysis requirements such as windowing, measuring execution time etc.
* Run the log cleaning command (remove the logs from the previous run of the pipeline)

# Step5. Run the pipeline

* Carefully submit the jobs to the cluster

# Step6. Post process

* Check the logs and if everything fine, remove them all