Software Testing

We report the testing methods below to verify both the artbin sample size and power results.

Tests 1-8 can be run all together from the main testing file: main\_testing\_file.do

All of these files are available on GitHub: [https://github.com/UCL/artbin](https://github.com/UCL/artbin%20%20%20)

|  |  |  |  |
| --- | --- | --- | --- |
| Corresponding number in SJ paper Software Testing Section | Testing Completed | Corresponding Stata testing files | Corresponding testing output files |
| 1. | We compared results for non-inferiority trials with those given by Julious and Owen (2011), Blackwelder (1982), Pocock (2003) and the online calculator  Sealed Envelope (2012). | artbin\_testing\_1.do | artbin\_testing\_1\_float.log  artbin\_testing\_1\_double.log  artbin comparison with published SS.docx |
| 2. | We compared results for a superiority binary outcome with those given by Pocock  (1983) and the online calculator Sealed Envelope (2012). | artbin\_testing\_2.do | artbin\_testing\_2\_float.log  artbin\_testing\_2\_double.log  artbin comparison with published SS.docx |
| 3. | We tested a range of scenarios including continuity correction results given by artbin  and those given by the Stata program power. | artbin\_testing\_3.do | artbin\_testing\_3\_float.log  artbin\_testing\_3\_double.log  artbin comparison with published SS.docx |
| 4. | We checked the results given by artbin using the margin option against Julious and  Owen (2011). | artbin\_testing\_4.do | artbin\_testing\_4\_float.log  artbin\_testing\_4\_double.log  artbin comparison with published SS.docx |
| 5. | The output of artbin was compared to Cytel's software EAST which is a sophisticated  package able to produce sample size and power calculations for a range of  binary outcomes in clinical trial settings. | artbin\_testing\_5.do | artbin\_testing\_5.log  artbin comparison with EAST.docx |
| 6. | For the new 'switch-on' syntax options we tested onesided for a one-sided test and  ccorrect to apply a continuity correction. | artbin\_testing\_6.do | artbin\_testing\_6\_float.log  artbin\_testing\_6\_double.log |
| 7. | We tested every permutation of 2-arm/more than 2-arms and non-inferiority/substantial superiority/  superiority trials with margin, local/distant, conditional/unconditional,  trend and Wald test options to check that the results were as expected, and that sample size was increased/decreased accordingly. | artbin\_testing\_7.do  explore\_every\_option.do (IW)  explore\_artbin.do (IW) | artbin\_testing\_7.log  explore\_every\_option.log (IW)  explore\_artbin.log (IW)  explore\_artbin\_postfile.dta (IW) |
| 8. | We checked error messages in a number of impossible cases, to ensure that we  obtained error messages as required. | artbin\_errortest\_8.do | artbin\_errortest\_8.log |
| 9. | We tested the dialogue box menu options to verify that the results were as required. | artbin\_dlgboxtesting\_9.do | artbin\_dlgboxoutput\_9.txt |