

1 **Quality in Clinical Research: An Observational**  
2 **Study of Randomisation Techniques in**  
3 **Urological and General Surgical Studies**  
4 **Short Title: Quality of Randomisation Techniques in Urological and**  
5 **General Surgical Studies**

6  
7 **Authors**

8 Nicholas Raison<sup>1,2</sup>, Simone Giona<sup>3</sup>, Oliver Brunckhorst<sup>1</sup>, Alexander Cohen<sup>4</sup>, Gordon  
9 Muir<sup>5</sup>

10

11 <sup>1</sup>MRC Centre for Transplantation, King's College London, London, UK

12 <sup>2</sup>Department of Urology, Guy's and St Thomas' NHS Trust, London, UK

13 <sup>3</sup>Department of Urology, Frimley Park Hospital, Camberley, UK

14 <sup>4</sup>Department of Haematological Medicine, Guy's and St Thomas' Hospitals NHS  
15 Foundation Trust, London, UK

16 <sup>5</sup>Department of Urology, King's College Hospital, London, UK

17

18 Correspondence to Mr G Muir, Consultant Urologist, Department of Urology King's

19 College Hospital, Denmark Hill, London, SE5 9RS

20 Email: [gordonhmuir@gmail.com](mailto:gordonhmuir@gmail.com)

21 Tele: +44 (0)20 3299 9000

22

23 Manuscript word Count: 2539 words

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

## 24 **ABSTRACT**

25

### 26 **Objectives**

27 To quantitatively test the quality of randomised controlled trials reported to  
28 international scientific meetings through a critical analysis of randomisation  
29 outcomes.

30

### 31 **Design and Main Outcome Measures**

32 All randomised controlled trials presented at international surgical and urological  
33 congresses using simple randomisation were identified. Primary analysis of  
34 randomisation technique was performed by comparing the observed and expected  
35 numbers of trials with equal numbers of participants in each arm. Sensitivity  
36 analyses compared study design, type of study and presence of external  
37 sponsorship. All abstracts were assessed according to the CONSORT for reporting  
38 randomised trials in journal and conference abstracts checklist.

39

### 40 **Results**

41 345 studies met the inclusion criteria. 148 studies reported simple randomisation to  
42 allocate 26,510 patients. Randomisation technique could not be identified in 104  
43 studies. Primary and all secondary analyses demonstrated a probability of  $p < 0.0001$   
44 that simple randomisation was used for participant allocation in all studies. Mean  
45 consort score was 9.4

46

### 47 **Conclusions**

48 It is extremely unlikely that simple randomisation was performed as reported in a  
49 significant proportion of the 148 RCTs in this study. These results raise concerning  
50 questions of the veracity and reliability of current medical research. There needs to  
51 be a greater awareness of the potential for methodological inaccuracy and error.

52

53 Key Words: Randomised Controlled Trial, Research, Quality, CONSORT, Trials

54 Methodology, Academic Surgery

55

56

## 57 INTRODUCTION

58

59 The randomised controlled trial (RCT) has long been considered the non-plus ultra of  
60 empirical research. It offers the most rigorous tool for researchers investigating the  
61 cause and effect relationship of any given collection of factors. The challenge was  
62 first set by Archie Cochrane who argued in 1972 that no new treatments should be  
63 introduced until their utility over existing treatments had been proven in a RCT(1).

64

65 Effective RCTs rely on the equal distribution of confounding factors between groups,  
66 cancelling out their effects and leaving variations in outcomes attributable only to  
67 chance or the intervention. Unbiased allocation is essential to the statistical analysis  
68 used to test for significance. RCTs are analysed on the assumption that any  
69 differences between groups will behave like differences between random samples of  
70 a single population. Importantly successful randomisation minimises the influence of  
71 investigators by ensuring that treatments or interventions cannot be predicted. The  
72 risk of substantial allocation bias, both conscious and unconscious, when  
73 investigators knowing or can predict the participants' allocation is well recognised  
74 with outcomes exaggerated by up to 41%(2).

75

76 RCTs are not immune to fraud or deceit, particularly if not subject to overview by a  
77 third party. The estimated prevalence of fraudulent behaviour in medical research  
78 varies enormously from less than one per cent to over 25%(3). Risks from weak or  
79 erroneous trial methodologies introducing potentially profound bias into studies are  
80 thankfully now widely acknowledged. The Consolidated Statement of Reporting  
81 Trials (CONSORT) statement provides a clear minimum checklist of items that

82 should be reported by RCTs. This has helped both investigators and readers assess  
83 the validity of studies against standard criteria. Numerous reviews have  
84 subsequently appraised various collections of studies against these standards(2, 4-  
85 7). These studies highlight the inadequate level of understanding of trial  
86 methodology which persists throughout medicine. Reporting of randomisation and  
87 allocation methods has been noted to be particularly poor(5, 6). Yet such studies  
88 remain limited by the lack of any objective assessment of methodological quality,  
89 relying instead on self-reporting by authors. The current study aims to objectively  
90 assess the quality (and potentially the veracity) of randomised controlled trials  
91 through the analysis of allocation outcomes following simple randomisation.

92

### 93 **Materials and Methods**

94

95 The study was conducted in accordance with STROBE reporting guidelines(8). All  
96 poster presentations of surgical RCTs presented at nine international surgical  
97 conferences covering urology and general surgery between 2012 and 2017 were  
98 included in the study (full list in supplementary table 1). Two authors (NR, SG)  
99 independently identified studies before each conference through hand searches of  
100 abstract books. Studies were included if they identified themselves as randomised  
101 controlled trials in the title or abstract or reported the random allocation of patients to  
102 study arms. Studies were excluded if they did not report original data or re-analysed  
103 previously collected data; if details of randomisation or allocation were not available;  
104 if trials did not involve the randomisation of human participants or if the trial was  
105 ongoing.

106

107 A stepwise approach to data collection was undertaken. Abstracts were reviewed for  
108 details of the randomisation technique together with the numbers of patients  
109 allocated to each study arm. The CONSORT definition of simple randomisation was  
110 used. Where possible, during each conference, presentations were attended and the  
111 study authors were directly questioned as to the randomisation technique and  
112 allocation outcomes. For those cases where authors were unavailable or did not  
113 attend, they were contacted directly by email (initially the presenting author then all  
114 study authors for whom contact details were available). If no email address or  
115 contact details were provided with the poster presentation, internet searches were  
116 undertaken to find email addresses. If no email address was found or there was no  
117 response, subsequent full publications were reviewed to identify randomisation  
118 technique. All authors were then contacted to confirm the same study design was  
119 used in the abstract presented and the paper published. Studies in which  
120 randomisation details were not confirmed by any such method were excluded.

121

122 A standardised data collection template was prepared prior to data collection. Data  
123 were collected on randomisation technique; number of study arms and number of  
124 participants allocated to each arm; multicentre or single centre trial design; presence  
125 of external funding; prospective trial registration; type of trial (medical/ surgical/  
126 mixed). “Medical trials” were defined as those assessing the use of a  
127 pharmacological agent or drug. “Surgical trials” were defined as any involving the  
128 direct assessment of a surgical intervention. Mixed trials were those that either  
129 combined a medical and surgical intervention, used a non-surgical intervention such  
130 as a catheter or involved a diagnostic intervention for example magnetic resonance  
131 imaging. All abstracts were assessed and scored according to the CONSORT for

132 reporting randomised trials in journal and conference abstracts checklist(9). Inclusion  
133 into all public clinical trial registries was manually checked for all studies.

134

### 135 **Statistical Analysis**

136

137 The analysis of randomisation techniques was performed by comparing the expected  
138 number of balanced studies with the observed number of balanced studies. A study  
139 was assumed to be balanced if the difference between the number of patients in  
140 each trial arm(s) was 1 or 0. Studies with unequal randomisation ratios (e.g. 2:1)  
141 were normalised to the largest group of the trial (e.g. by multiplying the smallest  
142 group with 2). The expected number of balanced studies and a one-sided p-value  
143 were calculated to test if simple randomisation was effective in each study. A total of  
144 10,000 Monte Carlo simulations under the assumption of simple randomisation were  
145 run for each study. The expected number of balanced studies was calculated as the  
146 mean number of balanced studies per simulation cycle. One-sided p-values were  
147 calculated as the relative proportion of simulation cycles in which the number of  
148 balanced studies was equal or greater to the number of observed balanced studies.  
149 All simulations were performed using the Stata MP Version 14.2 (StataCorp, LP).

150

151 The primary analysis compared the observed and expected number of participants in  
152 all studies. Sensitivity analyses were performed by stratifying by study design, (single  
153 centre vs multicentre), type of study (medical/ surgical or mixed) and the presence of  
154 external sponsorship. Parametric analysis of consort scores was performed using an  
155 independent samples T-test or ANOVA as appropriate. These analyses were  
156 performed on SPSS Statistics version 25 (IBM Corp. Armonk, NY).

157

158 **Results**

159 During the nine conferences 504 RCTs were presented of which 345 met the  
160 inclusion criteria. A specific randomisation technique was identified in 241 studies.  
161 148 studies used simple randomisation to allocate 26,510 patients. 22 studies were  
162 classified as medical, 76 as surgical and 50 as mixed. The randomisation technique  
163 could not be identified by any method in 104 studies (Figure 1).

164

165 Primary outcome analysis and sensitivity analyses are shown in Table 1. Results for  
166 both the primary outcome measure and all subgroup analyses demonstrate a  
167 probability of  $p < 0.0001$  that participant allocation was based only on simple  
168 randomisation in all of the included studies. Observed numbers of balanced studies  
169 greatly exceeded expected numbers in all analyses.

170

171 Table 1: Results of primary analysis and sensitivity analyses

172

Study type		Expected number of balanced studies (mean)	Observed number of balanced studies (n)	Total number of studies (n)	One-sided p-value
All studies		14	90	148	<0.0001
Number of centres	Multi centre	2	10	25	<0.0001
	Single centre	12	80	123	<0.0001
Industry	No	10	74	119	<0.0001



sponsored?	Yes	3	16	29	<0.0001
Type of intervention	Medical	2	11	22	<0.0001
	Surgical	7	50	76	<0.0001
	Mixed	4	29	50	<0.0001

173

174

175

176 Mean CONSORT score for all studies was 9.4. Single centre trials had significantly  
177 lower CONSORT scores than multicentre trial designs (mean CONSORT score 9.2  
178 vs 10.3,  $p = 0.02$ ). There was no difference in consort scores neither between  
179 studies with and without industry funding (mean CONSORT score 9.9 vs 9.3  $p =$   
180 0.06) nor between medical, surgical and mixed studies (mean CONSORT score 9.8  
181 vs 9.3 vs 9.2,  $p = 0.31$ ). 28 studies (19%) were registered on a clinical trial registry.  
182 Registration was associated with a significantly higher CONSORT score (mean  
183 CONSORT score 10.5 vs 9.1,  $p < 0.0001$ ).

184

## 185 Discussion

186

187 Based on this analysis of 148 studies randomising 26,510, it is almost impossible  
188 that all studies used the simple randomisation protocol that was reported in the study  
189 methodology. No difference was seen on sensitivity analyses according to study  
190 design, presence of industrial funding or study type. Put into context, most clinical  
191 studies accept a one in twenty risk of a chance association as “significant” – here the  
192 chance is one in tens of thousands. This ignores the fact that almost one third of  
193 studies were excluded as we could not find any information on randomisation.

194

195 We also noted that a disproportionate number of studies had what we term “round  
196 numbers” of subjects (20, 50, 100, 200 etc) suggesting that no powering of the initial  
197 trial had taken place. So few reported any power calculations, and therefore may  
198 have been under powered or over powered, it was not possible to test this  
199 hypothesis.

200

201 RCTs, particularly in surgery, are complex, expensive and time consuming. Using an  
202 RCT to address a scientific or clinical questions is constrained by a number of  
203 factors. Hypotheses need to focus on relatively simple and discrete questions,  
204 frequently impractical in the clinical setting. Investigators must recruit enough  
205 participants to provide robust results, a difficult undertaken highlighted by the high  
206 numbers of underpowered studies(10). Finally, patients must be willing to undergo a  
207 randomly allocated intervention and exposure to a potentially inferior treatment must  
208 be ethically defended.

209

210 In the face of these challenges, it is therefore unsurprising that RCTs continue to  
211 suffer from poor methodologies. Systematic reviews have shown, in spite of the  
212 CONSORT guidelines, that RCTs continue to be published without critical details  
213 such as the randomisation technique(11). Reviews across various disciplines have  
214 shown consistently low rates of of adequate reporting of randomisation technique:  
215 dermatology 9% (n = 6), surgery 43% (n = 65), urology 28% (n = 42) and general  
216 medicine 19% (n= 463) (5, 11-13). The quality of abstract publications have been  
217 found to be even worse prompting the release of the CONSORT statement for  
218 abstracts,. Even in high impact journals such as the Lancet, BMJ and JAMA, only

219 0%, 7% and 3% of abstracts respectively reported randomisation techniques(11).

220 This study adds further evidence to support these concerning results. Despite  
221 extensive and focussed efforts by study authors, basic trial design details could not  
222 be identified in 30% of studies.

223

224 Alongside the primary outcome results demonstrating the extremely high likelihood  
225 that in many cases randomisation was performed incorrectly, numerous examples of  
226 a lack understanding were noted during interviews with presenting authors. These  
227 discussions were “off the record” and anonymous, so we cannot comment on any  
228 factors linked to those authors. A number of authors were not aware of the  
229 randomisation technique and two admitted that true randomisation was in fact not  
230 really performed! Other studies reported techniques such as patient preference,  
231 clinician availability and simply alternating between techniques. Quasi-randomisation  
232 methods such as according to the date of birth, hospital record number or day of the  
233 week were used in a number of studies. In all cases the studies reported using  
234 simple randomisation.

235

236 Interpretation of these results must be considered in the context of a number of  
237 limitations. We relied on the presenting authors to understand and accurately report  
238 the randomisation technique. In some cases, junior authors presenting the studies  
239 were clearly unfamiliar with the details of the trial methodology and may have  
240 described the randomisation technique incorrectly. Including only poster  
241 presentations rather than full publications in peer reviewed journals may also have  
242 led to the inclusion of a greater proportion of low-quality studies. However previous  
243 attempts by the current authors to gather randomisation data from authors of

244 published studies garnered response rates of less than 10%. As result the direct  
245 approach of questioning authors at their posters was felt to be justified. These  
246 results focus on urological studies as it is the specialty of the majority of the authors.  
247 As in previous work on CONSORT compliance, the study was extended to include  
248 general surgical and non-urological specialties(6). Although direct comparisons were  
249 not performed, results with high proportions of balanced studies were seen in both  
250 urological and non-urological studies. Likewise, author interviews revealed  
251 comparable numbers of suspect answers. Replication of this research in other  
252 disciplines would be welcome. Finally given the nature of the analysis undertaken in  
253 this study, it is only possible to report the overall likelihood that simple randomisation  
254 was performed as reported. No assessment can be made on the exact numbers of  
255 studies that reported erroneous methodologies.

256

257 This study included only RCTs that used simple randomisation. Simple  
258 randomisation represents the most straightforward approach to allocating patients  
259 but can still be difficult to implement (14). All participants should have an equal  
260 chance of being allocated to either arm. Especially with smaller sample sizes, this  
261 can result in unequal group sizes and important factors may be not be balanced  
262 between groups. Yet the false belief that samples sizes should be equal remains  
263 pervasive(7, 15). This misconception is reflected in our results with 59% of studies  
264 (n=88/148) reporting equal arms or differing by just 1.

265

266 Despite these drawbacks, RCT's remain a cornerstone for the advancement of  
267 medical practice. The COVID-19 pandemic has clearly shown the importance of high

268 quality RCTs with their results directing clinical management and policy across the  
269 world.

270

271 However often in everyday clinical practice, especially in surgery, such sophisticated  
272 study designs are potentially being overused. The value of well conducted, non-  
273 randomised studies is often overlooked by both researchers and funding bodies.

274 This study has shown that in many cases, RCT's are beyond the abilities of surgical  
275 teams. Sadly medical history is littered with poorly executed and fraudulent studies  
276 whose biased results have misled medical opinion, occasionally for decades.

277 Alarming figures for the prevalence of research misconduct and rising numbers of  
278 retractions have prompted many national government to appoint national bodies to  
279 regulate and provide oversight on academic research(16). The results from this  
280 study suggest similar practises continue in academic research. Whilst a proportion of  
281 errors will be attributable to a lack of experience and ability of the study authors,  
282 wilful dishonesty must also be considered as factor. The common conclusion to so  
283 many non-randomised studies that an RCT is required to provide the definitive  
284 answer must be reviewed. Instead the utility of well-designed comparative studies  
285 should be better advocated. For the majority of smaller studies, a non-randomised  
286 trial design would help greatly simplify the studies, enabling surgical groups focus on  
287 achieving true and reliable results.

288

## 289 **Conclusion**

290

291 This study has shown that it is extremely likely that for a significant proportion of 148  
292 simple randomised studies, randomisation was not performed as described. All

293 studies had been selected for presentation following a peer review process and their  
294 results reported in international scientific journals. These results raise important  
295 questions on the current quality of medical research and reliability of clinicians to  
296 undertake research.

297

298 We cannot say why these results have occurred – ignorance of the RCT process  
299 may be the reason, yet we know from personal experience that the majority of  
300 institutions presenting data have taken part in high quality multicentre RCT's which  
301 would not fall foul of our scrutiny. Pressure to publish may mean that, particularly for  
302 single centres studies, corners were cut and proper protocols were ignored for  
303 simplicity. All trials must be registered prospectively with clear, publicly disseminated  
304 randomisation protocols. Of course, fraud is an ever-present explanation for  
305 research findings which are either too good to be true or which do not make sense.  
306 Faced with the above however, it would seem that any single centre RCT submitted  
307 is likely to be greatly at risk of basic statistical flaws – we would suggest that editors  
308 would be well advised to seek a statistician's advice before sending such trials out to  
309 hard pressed reviewers.

310

311 **Competing Interests:** For all authors none are declared

312 **Source of Funding:** For all authors none are declared

313 **Ethical approval:** Ethical approval not required.

314 **Contributorship Statement:** All authors meet the criteria for authorship

315 **Guarantor:** The guarantor, Mr G Muir, takes responsibility for the integrity of the  
316 work and confirms that he controlled the decision to publish.

317 **Acknowledgements:** We thank Stephen Rietbrock and Christopher of the  
318 Wallenhorst, Institute for Epidemiology, Statistics and Informatics for performing the  
319 statistical analysis. Neither were received compensation for their role in the study  
320

## 321 **References**

322

- 323 1. Cochrane AL. Effectiveness and efficiency: random reflections on health services.  
324 London: Nuffield Provincial Hospitals Trust; 1972. xi, 92 p. p.
- 325 2. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions  
326 of methodological quality associated with estimates of treatment effects in controlled trials.  
327 JAMA. 1995;273(5):408-12.
- 328 3. Sarwar U, Nicolaou M. Fraud and deceit in medical research. J Res Med Sci.  
329 2012;17(11):1077-81.
- 330 4. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated  
331 standards of reporting trials (CONSORT) and the completeness of reporting of randomised  
332 controlled trials (RCTs) published in medical journals. Cochrane Database Syst Rev.  
333 2012;11:MR000030.
- 334 5. Adie S, Harris IA, Naylor JM, Mittal R. CONSORT compliance in surgical randomized  
335 trials: are we there yet? A systematic review. Ann Surg. 2013;258(6):872-8.
- 336 6. Agha R, Cooper D, Muir G. The reporting quality of randomised controlled trials in  
337 surgery: a systematic review. Int J Surg. 2007;5(6):413-22.
- 338 7. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization  
339 from reports of controlled trials published in obstetrics and gynecology journals. JAMA.  
340 1994;272(2):125-8.

- 341 8. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al.  
342 The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)  
343 statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61(4):344-  
344 9.
- 345 9. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT  
346 for reporting randomised trials in journal and conference abstracts. *Lancet*.  
347 2008;371(9609):281-3.
- 348 10. Dimick JB, Diener-West M, Lipsett PA. Negative results of randomized clinical trials  
349 published in the surgical literature: equivalency or error? *Arch Surg*. 2001;136(7):796-800.
- 350 11. Hays M, Andrews M, Wilson R, Callender D, O'Malley PG, Douglas K. Reporting  
351 quality of randomised controlled trial abstracts among high-impact general medical  
352 journals: a review and analysis. *BMJ Open*. 2016;6(7):e011082.
- 353 12. Scales CD, Jr., Norris RD, Keitz SA, Peterson BL, Preminger GM, Vieweg J, et al. A  
354 critical assessment of the quality of reporting of randomized, controlled trials in the urology  
355 literature. *J Urol*. 2007;177(3):1090-4; discussion 4-5.
- 356 13. Adetugbo K, Williams H. How well are randomized controlled trials reported in the  
357 dermatology literature? *Arch Dermatol*. 2000;136(3):381-5.
- 358 14. Kao LS, Tyson JE, Blakely ML, Lally KP. Clinical research methodology I: introduction  
359 to randomized trials. *J Am Coll Surg*. 2008;206(2):361-9.
- 360 15. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials.  
361 *Lancet*. 1990;335(8682):149-53.
- 362 16. Tavaré A. Managing research misconduct: is anyone getting it right? *BMJ*.  
363 2011;343:d8212.
- 364



365 **Figure Legends**

366 Figure 1: Study Flow Diagram

367

368 **Supplementary Table Legends**

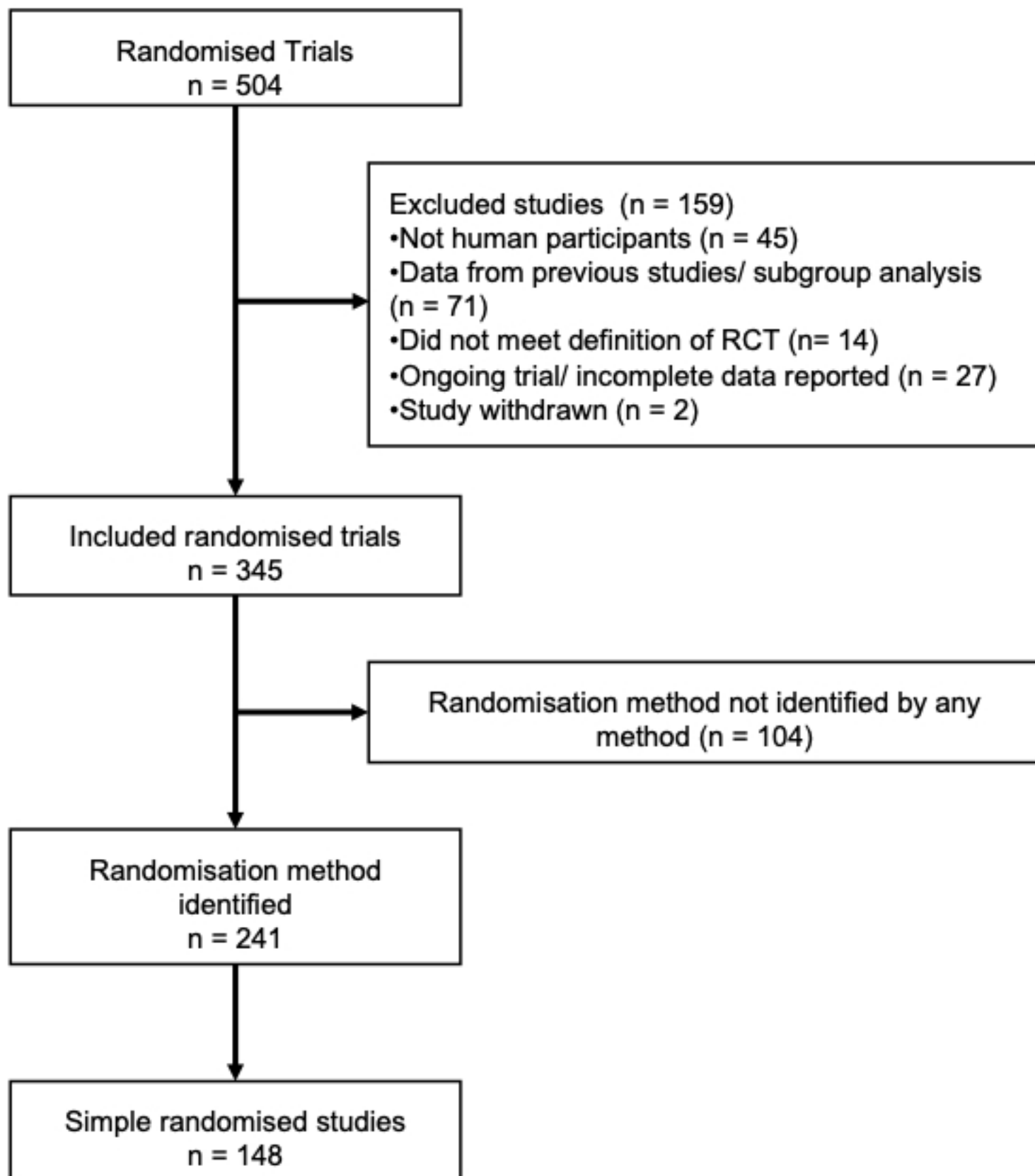
369 Supplementary Table 1: Surgical and Urological Conferences Included in this Study

370

371

372

## Study Flow Diagram



Figure