# Novel statistical investigation methods examining data integrity for 33 randomized trials in 18 journals from one research group

**‘The empty vessel makes the greatest sound’**
**Shakespeare, *Henry V***

**Mark Bolland, University of Auckland**
**Alison Avenell*, University of Aberdeen**
**Andrew Grey, University of Auckland**
**Greg Gamble, University of Auckland**

# Our presentation in three acts

I **Novel statistical investigation methods examining data integrity for 33 randomized trials in 18 journals from one research group**

II **Investigating the impact of retracted randomized clinical trial reports**

III **Reporting concerns about data integrity for 33 randomized trials in 18 journals from one research group: a narrative review**

# Conflict of interest statements

- None of the authors has a conflict to disclose
- All authors wish to improve the integrity of the research literature

- **Analysing concerns about data integrity**
  - **P-values for comparisons of baseline groups**
  - **Distribution of standardized sample means for baseline continuous variables**
  - **Assessing randomization**
  - **Likelihood of outcome results**

**Appendix e-1:**

**Table e-1A: 33 Randomized controlled trials carried out by the researchers.**

1. Sato Y, Maruoka H, Oizumi K. Amelioration of hemiplegia-associated osteopenia more than 4 years after stroke by 1 alpha-hydroxyvitamin D3 and calcium supplementation. *Stroke* 1997;28:736-9.
2. Sato Y, Honda Y, Kuno H, Oizumi K. Menatetrenone ameliorates osteopenia in disuse-affected limbs of vitamin D- and K-deficient stroke patients. *Bone* 1998;23:291-6.
3. Sato Y, Kuno H, Kaji M, Saruwatari N, Oizumi K. Effect of ipriflavone on bone in elderly hemiplegic stroke patients with hypovitaminosis D. *Am J Phys Med Rehabil* 1999;78:457-63.
4. Sato Y, Manabe S, Kuno H, Oizumi K. Amelioration of osteopenia and hypovitaminosis D by 1alpha-hydroxyvitamin D3 in elderly patients with Parkinson's disease. *J Neurol Neurosurg Psychiatry* 1999;66:64-8.
5. Sato Y, Asoh T, Kaji M, Oizumi K. Beneficial effect of intermittent cyclical etidronate therapy in hemiplegic patients following an acute stroke. *J Bone Miner Res* 2000;15:2487-94.
6. Sato Y, Honda Y, Kaji M, Asoh T, Hosokawa K, Kondo I, et al. Amelioration of osteoporosis by menatetrenone in elderly female Parkinson's disease patients with vitamin D deficiency. *Bone* 2002;31:114-8.
7. Sato Y, Kaji M, Kondo I, Yoshida H, Satoh K, Metoki N. Hyperhomocysteinemia in Japanese patients with convalescent stage ischemic stroke: effect of combined therapy with folic acid and mecobalamine. *J Neurol Sci* 2002;202:65-8.
8. Sato Y, Asoh T, Metoki N, Satoh K. Efficacy of methylprednisolone pulse therapy on neuroleptic malignant syndrome in Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2003;74:574-6.
9. Sato Y, Metoki N, Iwamoto J, Satoh K. Amelioration of osteoporosis and hypovitaminosis D by sunlight exposure in stroke patients. *Neurology* 2003;61:338-42.
10. Sato Y, Kanoko T, Yasuda H, Satoh K, Iwamoto J. Beneficial effect of etidronate therapy in immobilized hip fracture patients. *Am J Phys Med Rehabil* 2004;83:298-303.
11. Iwamoto J, Takeda T, Sato Y, Uzawa M. Comparison of effect of treatment with etidronate and alendronate on lumbar bone mineral density in elderly women with osteoporosis. *Yonsei Med J* 2005;46:750-8.
12. Sato Y, Honda Y, Iwamoto J, Kanoko T, Satoh K. Amelioration by mecobalamin of subclinical carpal tunnel syndrome involving unaffected limbs in stroke patients. *J Neurol Sci* 2005;231:13-8.
13. Sato Y, Honda Y, Iwamoto J, Kanoko T, Satoh K. Effect of folate and mecobalamin on hip f patients with stroke: a randomized controlled trial. *JAMA* 2005;293:1082-8.
14. Sato Y, Iwamoto J, Kanoko T, Satoh K. Low-dose vitamin D prevents muscular atrophy and and hip fractures in women after stroke: a randomized controlled trial. *Cerebrovasc Dis* 2005
15. Sato Y, Iwamoto J, Kanoko T, Satoh K. Risedronate sodium therapy for prevention of hip fra 65 years or older after stroke. *Arch Intern Med* 2005;165:1743-8.
16. Sato Y, Iwamoto J, Kanoko T, Satoh K. Amelioration of osteoporosis and hypovitaminosis D exposure in hospitalized, elderly women with Alzheimer's disease: a randomized controlled t *Miner Res* 2005;20:1327-33.
17. Sato Y, Iwamoto J, Kanoko T, Satoh K. Risedronate therapy for prevention of hip fracture af elderly women. *Neurology* 2005;64:811-6.
18. Sato Y, Kanoko T, Satoh K, Iwamoto J. The prevention of hip fracture with risedronate and e plus calcium supplementation in elderly women with Alzheimer disease: a randomized contr *Arch Intern Med* 2005;165:1737-42.
19. Sato Y, Kanoko T, Satoh K, Iwamoto J. Menatetrenone and vitamin D2 with calcium supplei nonvertebral fracture in elderly women with Alzheimer's disease. *Bone* 2005;36:61-8.
20. Iwamoto J, Takeda T, Sato Y, Uzawa M. Effect of whole-body vibration exercise on lumbar density, bone turnover, and chronic back pain in post-menopausal osteoporotic women treate alendronate. *Aging Clin Exp Res* 2005;17:157-63.
21. Sato Y, Honda Y, Iwamoto J. Etidronate for fracture prevention in amyotrophic lateral sclero randomized controlled trial. *Bone* 2006;39:1080-6.
22. Sato Y, Iwamoto J, Kanoko T, Satoh K. Alendronate and vitamin D2 for prevention of hip fr Parkinson's disease: a randomized controlled trial. *Mov Disord* 2006;21:924-9.
23. Sato Y, Honda Y, Iwamoto J. Risedronate and ergocalciferol prevent hip fracture in elderly Parkinson disease. *Neurology* 2007;68:911-5.
24. Iwamoto J, Sato Y, Uzawa M, Takeda T, Matsumoto H. Comparison of effects of alendronate and raloxifene on lumbar bone mineral density, bone turnover, and lipid metabolism in elderly women with osteoporosis. *Yonsei Med J* 2008;49:119-28.
25. Iwamoto J, Sato Y, Uzawa M, Takeda T, Matsumoto H. Comparison of the effects of alendronate and alfacalcidol on hip bone mineral density and bone turnover in Japanese men having osteoporosis or osteopenia with clinical risk factors for fractures. *Yonsei Med J* 2009;50:474-81.
26. Iwamoto J, Suzuki H, Tanaka K, Kumakubo T, Hirabayashi H, Miyazaki Y, et al. Preventative effect of exercise against falls in the elderly: a randomized controlled trial. *Osteoporos Int* 2009;20:1233-40.
27. Sato Y, Iwamoto J, Honda Y. Beneficial effect of etidronate therapy in chronically hospitalized, disabled patients with stroke. *J Stroke Cerebrovasc Dis* 2010;19:198-203.
28. Iwamoto J, Makita K, Sato Y, Takeda T, Matsumoto H. Alendronate is more effective than elcatonin in improving pain and quality of life in postmenopausal women with osteoporosis. *Osteoporos Int* 2011;22:2735-42.
29. Sato Y, Honda Y, Umeno K, Hayashida N, Iwamoto J, Takeda T, et al. The prevention of hip fracture with menatetrenone and risedronate plus calcium supplementation in elderly patients with Alzheimer disease: a randomized controlled trial. *Kurume Med J* 2011;57:117-24.
30. Sato Y, Iwamoto J, Honda Y. Once-weekly risedronate for prevention of hip fracture in women with Parkinson's disease: a randomised controlled trial. *J Neurol Neurosurg Psychiatry* 2011;82:1390-3.
31. Sato Y, Iwamoto J, Honda Y. Amelioration of osteoporosis and hypovitaminosis D by sunlight exposure in Parkinson's disease. *Parkinsonism Relat Disord* 2011;17:22-6.
32. Sato Y, Iwamoto J, Honda Y. An open-label trial comparing alendronate and alphacalcidol in reducing falls and hip fractures in disabled stroke patients. *J Stroke Cerebrovasc Dis* 2011;20:41-6.
33. Iwamoto J, Sato Y, Takeda T, Matsumoto H. Whole body vibration exercise improves body balance and walking velocity in postmenopausal osteoporotic women treated with alendronate: Galileo and Alendronate Intervention Trail (GAIT). *J Musculoskelet Neuronal Interact* 2012;12:136-43.



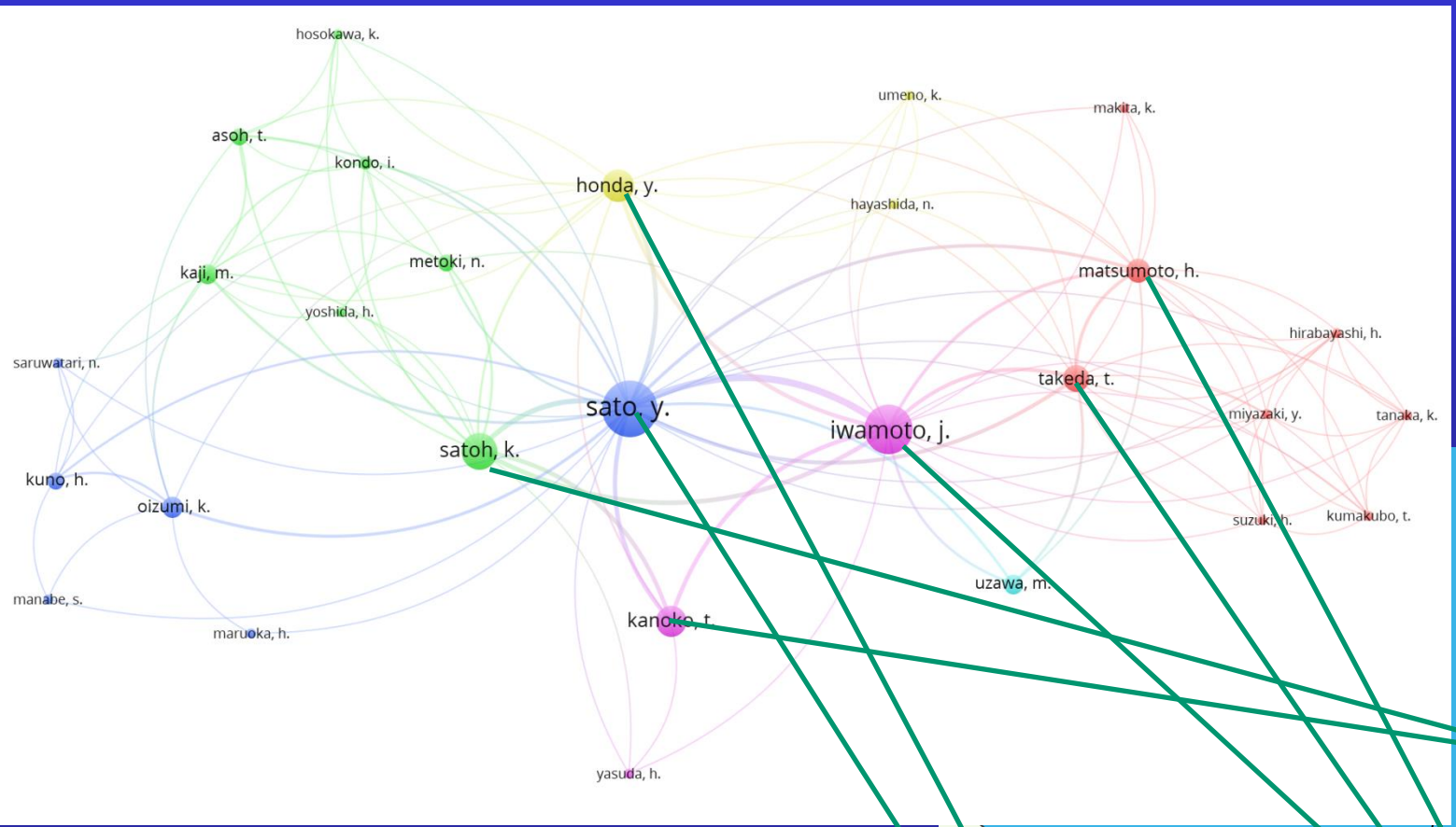
**33 RCTs**

**1997-2012**

**N = 6253**

**26 Authors**

**12 Institutions**




**33 RCTs from >220 publications from the two lead authors**


**Y Sato, Neurologist**
**Mitate Hospital**
**410 bed hospital, 8 Drs, 102 nurses**

**J Iwamoto, Sports medicine**
**Keio University, Tokyo**

# Remarkable productivity

- **Between March – July 2003 (5 months) recruited**
  - **500 ambulatory females, >70y with Alzheimer's disease living in the community in 2 months**
  - **280 males, >65y with hemiplegic stroke in 2 months**
  - **374 females, >65y with acute hemiplegic stroke in 4m**
  - **And**
    - **participants reviewed every four weeks**
    - **ongoing intensive follow-up in 3 other trials (n=774)**
    - **recruitment /intensive follow-up for 2 other trials (n=292)**
  - **But only same 4 co-authors on 4 trials, and 1 other author for 2 trials**

# Improbably similar randomized treatment groups- Part 1

**‘Baseline data indicate that the study groups are strikingly well matched. Given the relatively small study sizes, the high frequency of identical baseline variables appears unusual’**

**Response:**

**‘The high frequency of identical baseline variables was not by design but by chance. The group of neurological patients consist of homogeneous geriatric subjects, and their background characteristics fell into a relatively small range of variation.’**

**Halbekath et al. Arch Intern Med 2007;167:513-4**

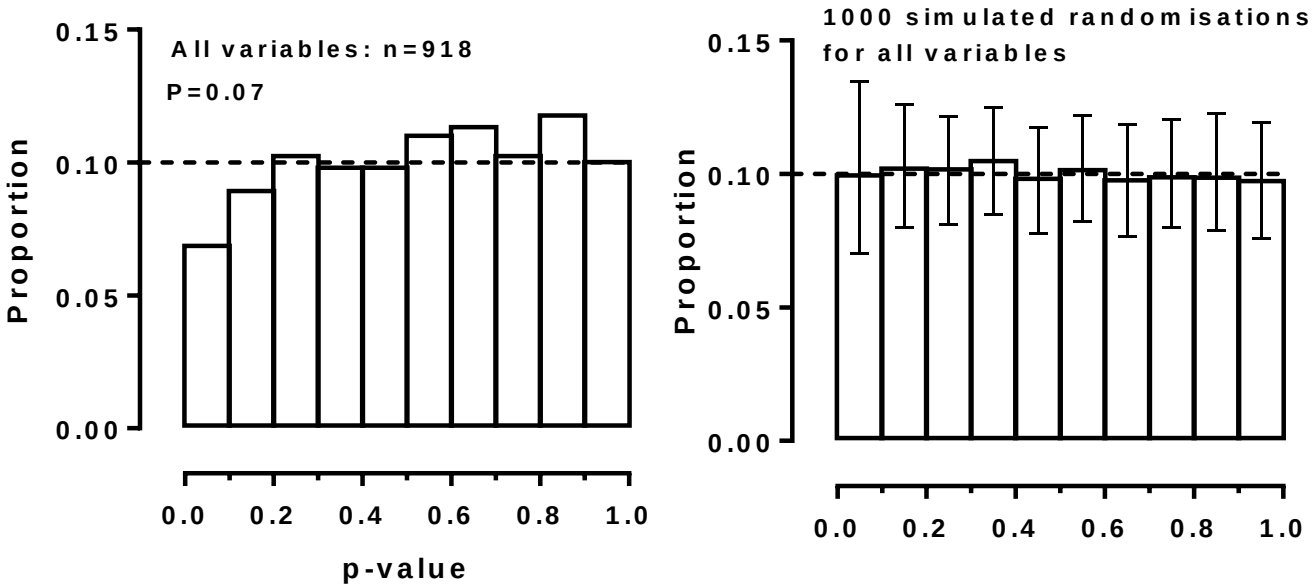
# Improbably similar randomized treatment groups- Part 1

- **Since allocation of participants in a RCT is random, comparisons between randomized groups for independent variables at baseline should produce a random distribution of p-values**

- **For example there is an equal likelihood of a p-value of <0.1 and >0.9, of <0.2 and >0.8, etc.**

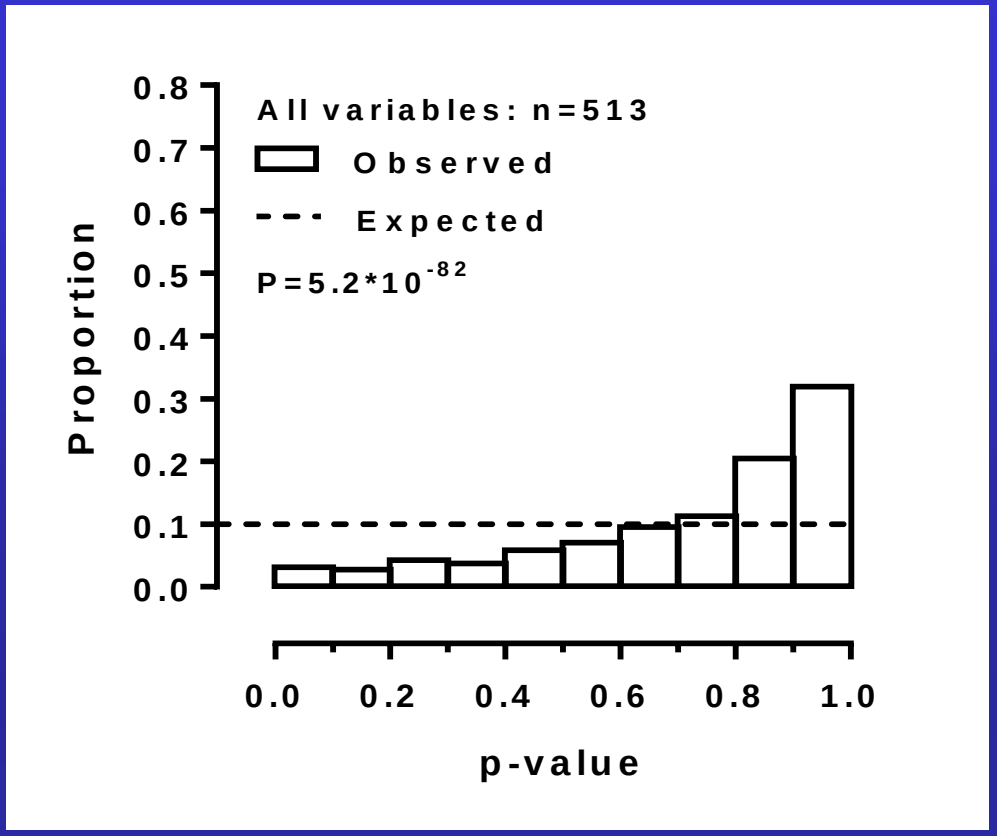
# Auckland RCTs

**Observed vs expected distribution of baseline p-values by decile**




**The dotted line shows the expected proportion and the error bars the 95% CI**

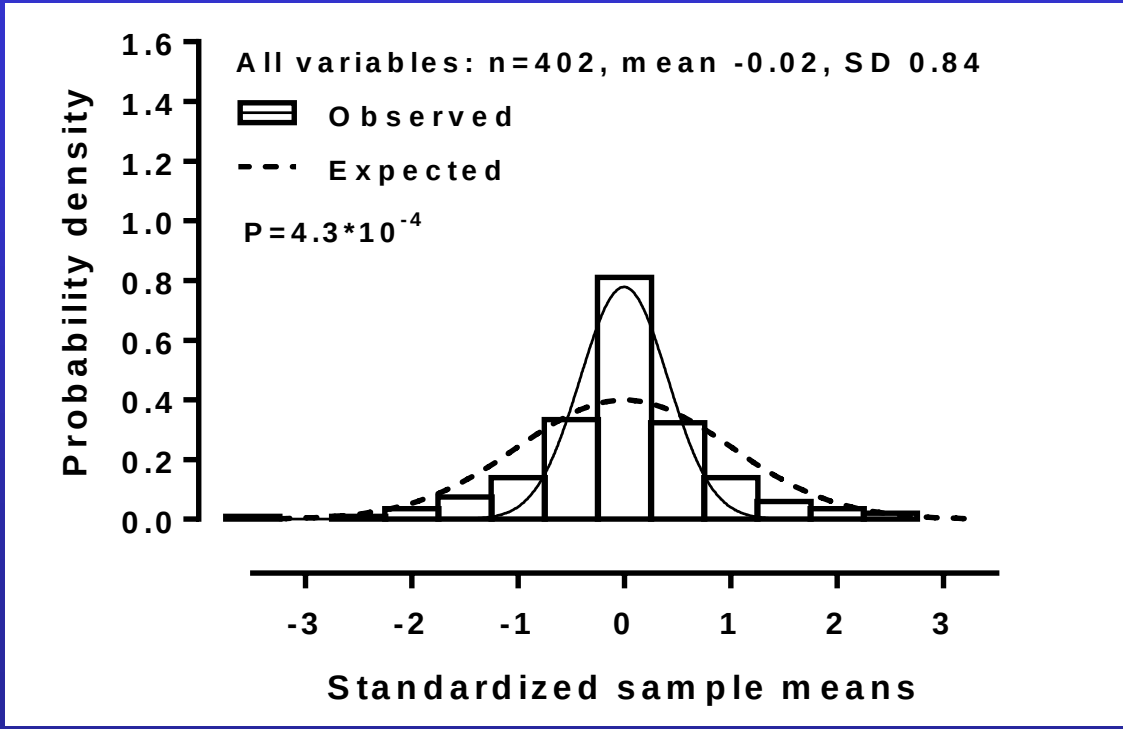
## 13 trials, 2851 participants




- **For 32 RCTs 513 baseline variables**
  - **52% of p-values were >0.8**
  - **6%, 14%, and 27% of p-values were <0.2, <0.4, and <0.6, respectively.**
  - **Highly unlikely to have arisen by chance ($P=5.2*10^{-82}$).**

# Improbably similar randomized treatment groups- Part 2

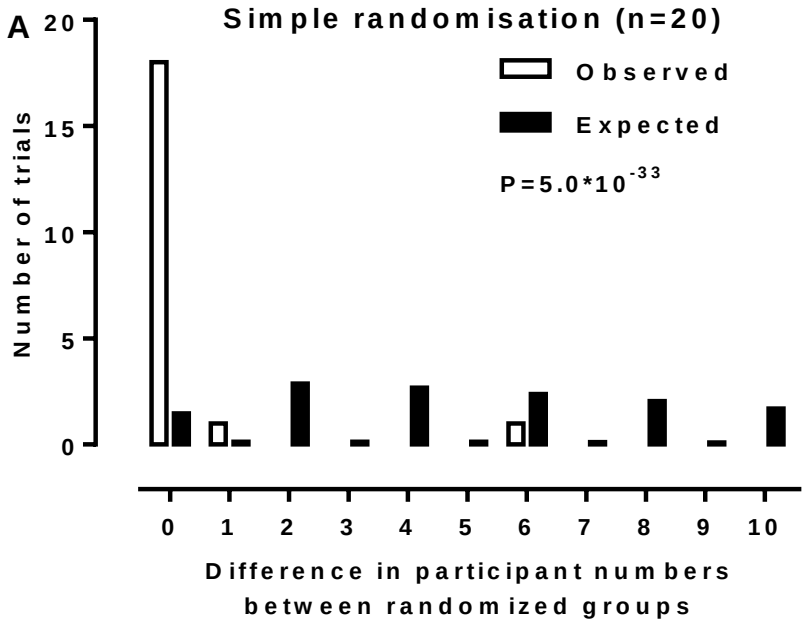
- **Similar principle to assess sampling distribution of continuous variables**
- **If population repeatedly sampled, means of samples are ~ normally distributed**
- **Can test whether distribution differs from expected**
- **Each randomized group represents a sample from the total study cohort**

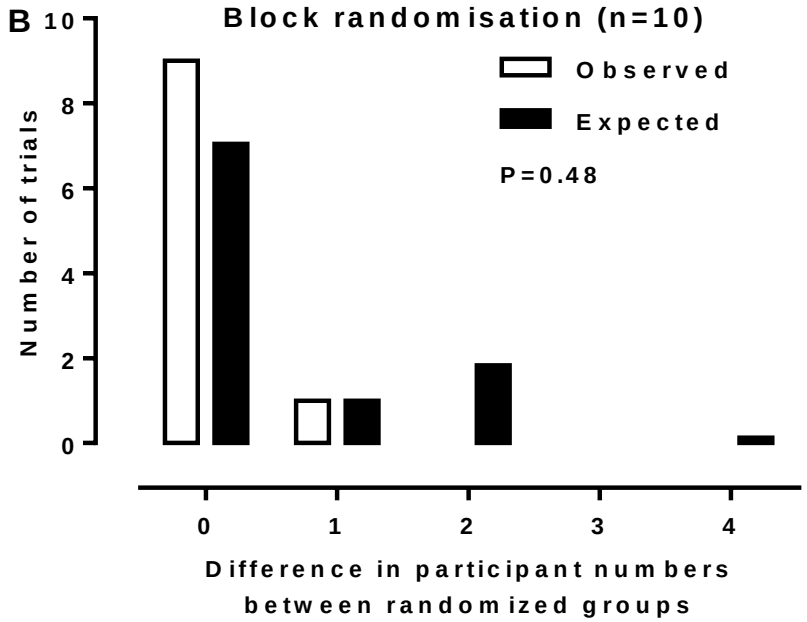



- **For 402 baseline continuous variables.**
  - **SD 0.84, distribution differs from expected ($P=4.3*10^{-4}$)**
  - **Values clustered more tightly around mean**

# Improbably similar randomized treatment groups- Part 3

- **In simple randomisation:**
  - **allocation of each participant is equivalent to coin toss**
  - **unlikely that groups will be same**
    - **N=10, 25% chance of two groups of 5; N=50, 11% chance of n=25 etc …**
- **In randomisation with small blocks:**
  - **differences between groups depends only on last block**
    - **If full - groups are same size**
    - **For block of 4:**
      - **If odd number in last block, groups must differ by 1**
      - **If last block = 2, 2/3 chance groups same size, 1/3 groups differ by 2**
- **For both situations:**
  - **Can calculate probability treatment groups are same, or differ by 0, 1, 2, 3, 4 etc for each trial…**
  - **Compare observed and expected distribution of differences**




- **27/30 2-arm RCTs same number of participants in each treatment group**
- **If we assumed <u>all</u> trials were randomized in blocks of 4:**
  - **Expected: 22 trials same, 5 trials differ by 2**
  - **Observed: 27 trials same, 0 differ by 2**

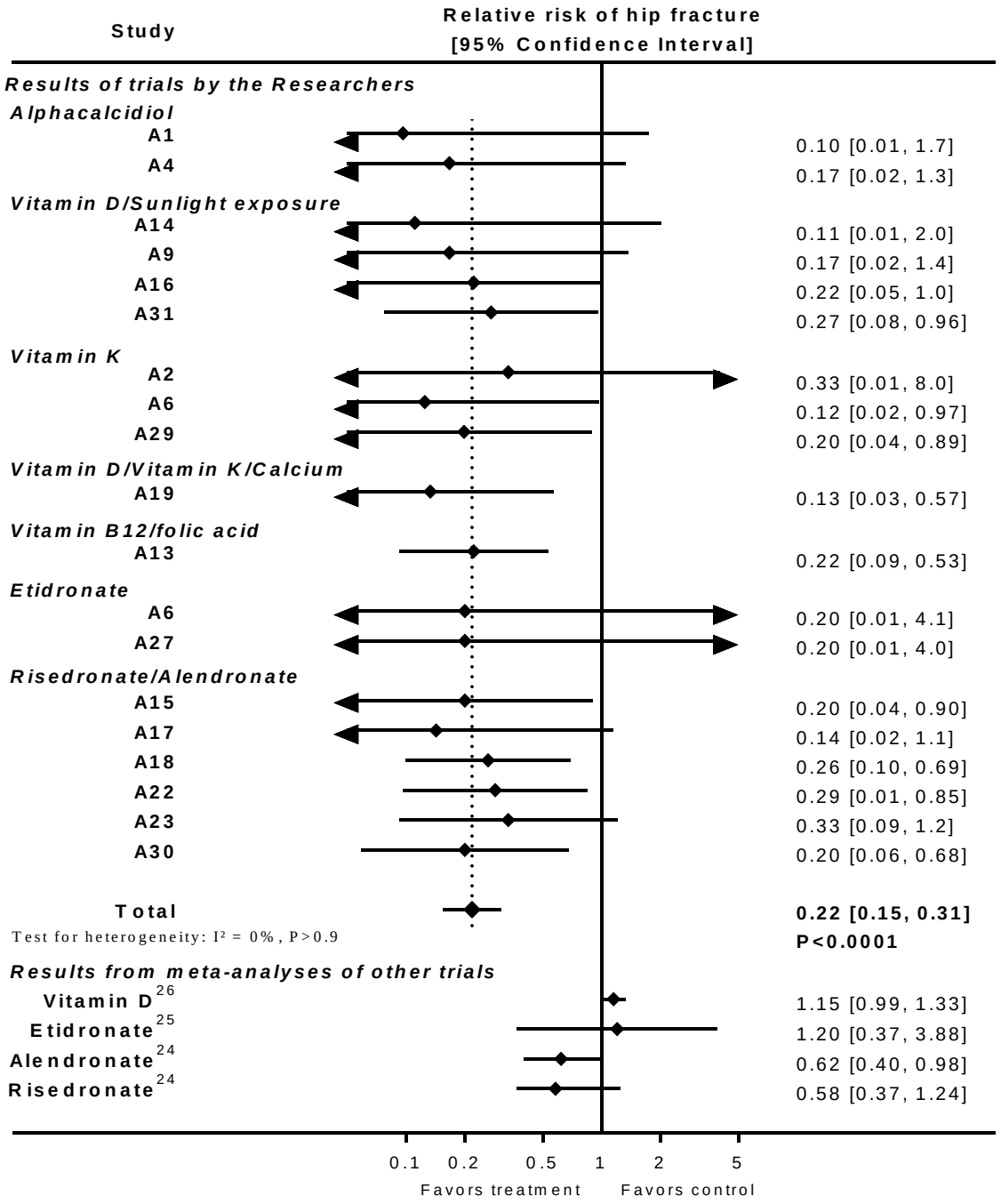
  **P = 0.04**

# Improbably similar randomized treatment groups

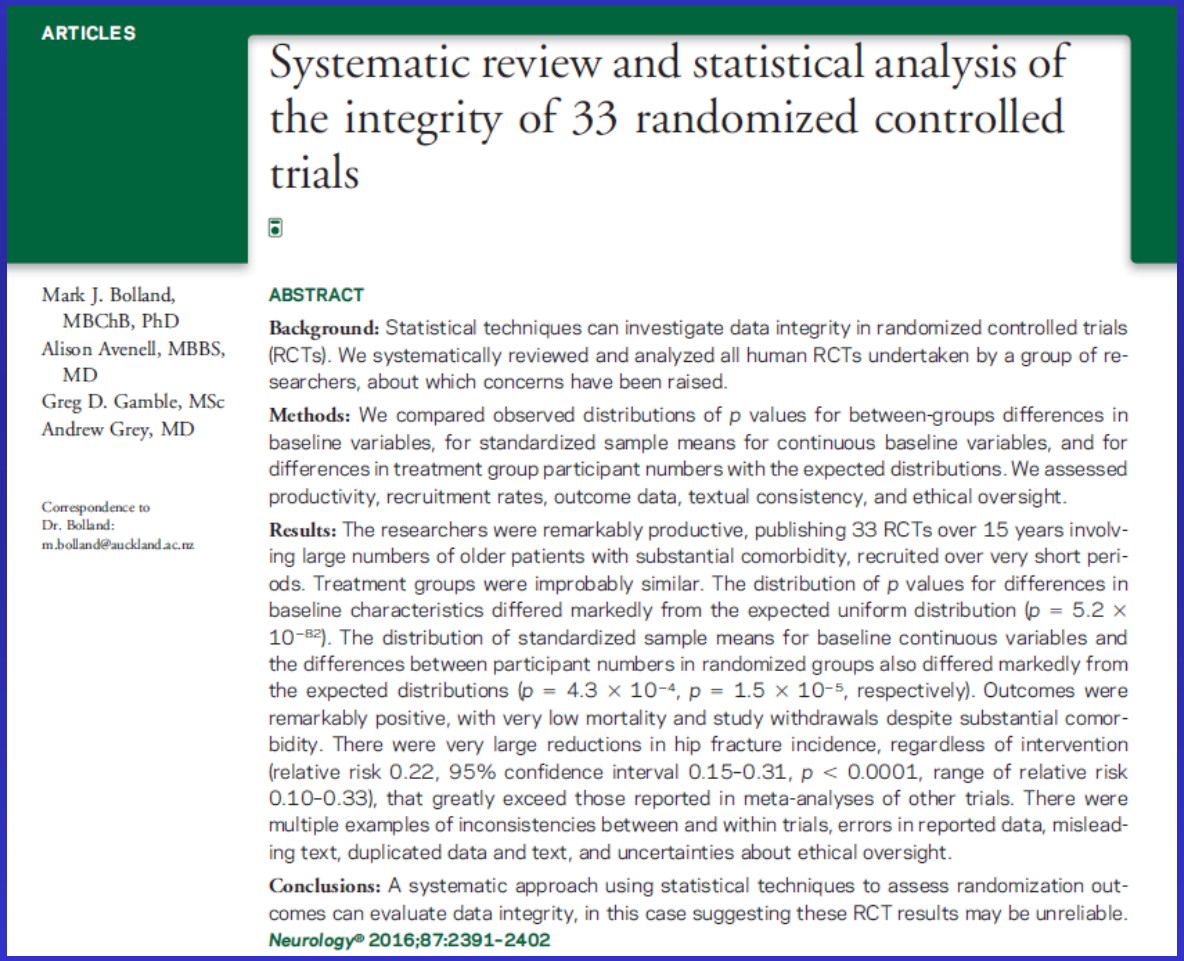
- **Strongly suggests studies were not randomised using simple randomisation or in small blocks**

# Remarkably positive outcomes

- **Despite studying frail elderly individuals with substantial co-morbidity, only 519 (8.8%) participants did not complete the trials**
  - **Surprising given very high rates of hip fracture and significant background co-morbidity, both of which are associated with high mortality.**
- **Strikingly consistent, large reductions in hip fracture risk from variety of agents, inconsistent with results from other trials**


Study
Relative risk of hip fracture
[95% Confidence Interval]
Results of trials by the Researchers
Alphacalcidiol
A1
0.10 [0.01, 1.7]
A4
0.17 [0.02, 1.3]
Vitamin D/Sunlight exposure
A14
0.11 [0.01, 2.0]
A9
0.17 [0.02, 1.4]
A16
0.22 [0.05, 1.0]
A31
0.27 [0.08, 0.96]
Vitamin K
A2
0.33 [0.01, 8.0]
A6
0.12 [0.02, 0.97]
A29
0.20 [0.04, 0.89]
Vitamin D/Vitamin K/Calcium
A19
0.13 [0.03, 0.57]
Vitamin B12/folic acid
A13
0.22 [0.09, 0.53]
Etidronate
A6
0.20 [0.01, 4.1]
A27
0.20 [0.01, 4.0]
Risedronate/Alendronate
A15
0.20 [0.04, 0.90]
A17
0.14 [0.02, 1.1]
A18
0.26 [0.10, 0.69]
A22
0.29 [0.01, 0.85]
A23
0.33 [0.09, 1.2]
A30
0.20 [0.06, 0.68]
Total
0.22 [0.15, 0.31]
Test for heterogeneity: I² = 0%, P>0.9
P<0.0001
Results from meta-analyses of other trials
Vitamin D[26]
1.15 [0.99, 1.33]
Etidronate[25]
1.20 [0.37, 3.88]
Alendronate[24]
0.62 [0.40, 0.98]
Risedronate[24]
0.58 [0.37, 1.24]
0.1
0.2
0.5
1
2
5
Favors treatment
Favors control

# Published in Neurology 2016; 87:2391-2402



ARTICLES

## Systematic review and statistical analysis of the integrity of 33 randomized controlled trials

Mark J. Bolland, MBChB, PhD
Alison Avenell, MBBS, MD
Greg D. Gamble, MSc
Andrew Grey, MD

Correspondence to Dr. Bolland: m.bolland@auckland.ac.nz

**ABSTRACT**

**Background:** Statistical techniques can investigate data integrity in randomized controlled trials (RCTs). We systematically reviewed and analyzed all human RCTs undertaken by a group of researchers, about which concerns have been raised.

**Methods:** We compared observed distributions of $p$ values for between-groups differences in baseline variables, for standardized sample means for continuous baseline variables, and for differences in treatment group participant numbers with the expected distributions. We assessed productivity, recruitment rates, outcome data, textual consistency, and ethical oversight.

**Results:** The researchers were remarkably productive, publishing 33 RCTs over 15 years involving large numbers of older patients with substantial comorbidity, recruited over very short periods. Treatment groups were improbably similar. The distribution of $p$ values for differences in baseline characteristics differed markedly from the expected uniform distribution ($p = 5.2 \times 10^{-82}$). The distribution of standardized sample means for baseline continuous variables and the differences between participant numbers in randomized groups also differed markedly from the expected distributions ($p = 4.3 \times 10^{-4}$, $p = 1.5 \times 10^{-5}$, respectively). Outcomes were remarkably positive, with very low mortality and study withdrawals despite substantial comorbidity. There were very large reductions in hip fracture incidence, regardless of intervention (relative risk 0.22, 95% confidence interval 0.15-0.31, $p < 0.0001$, range of relative risk 0.10-0.33), that greatly exceed those reported in meta-analyses of other trials. There were multiple examples of inconsistencies between and within trials, errors in reported data, misleading text, duplicated data and text, and uncertainties about ethical oversight.

**Conclusions:** A systematic approach using statistical techniques to assess randomization outcomes can evaluate data integrity, in this case suggesting these RCT results may be unreliable.

*Neurology*® 2016;87:2391-2402

# Conclusion

- **Serious concerns about integrity/validity of results**
- **Objective evidence is:**
  - **systematic failure of randomization**
  - **consistently outlying outcome data compared to other trials**
  - **remarkable productivity**
- **We also found**
  - **internal inconsistencies, duplication of data and text**
  - **numerous misleading statements and errors**
  - **concerns regarding ethical approval**

**Acknowledgements**

**Bill Gillespie (Hull York Medical School, UK)**

**Lesley Gillespie (University of Otago, New Zealand)**

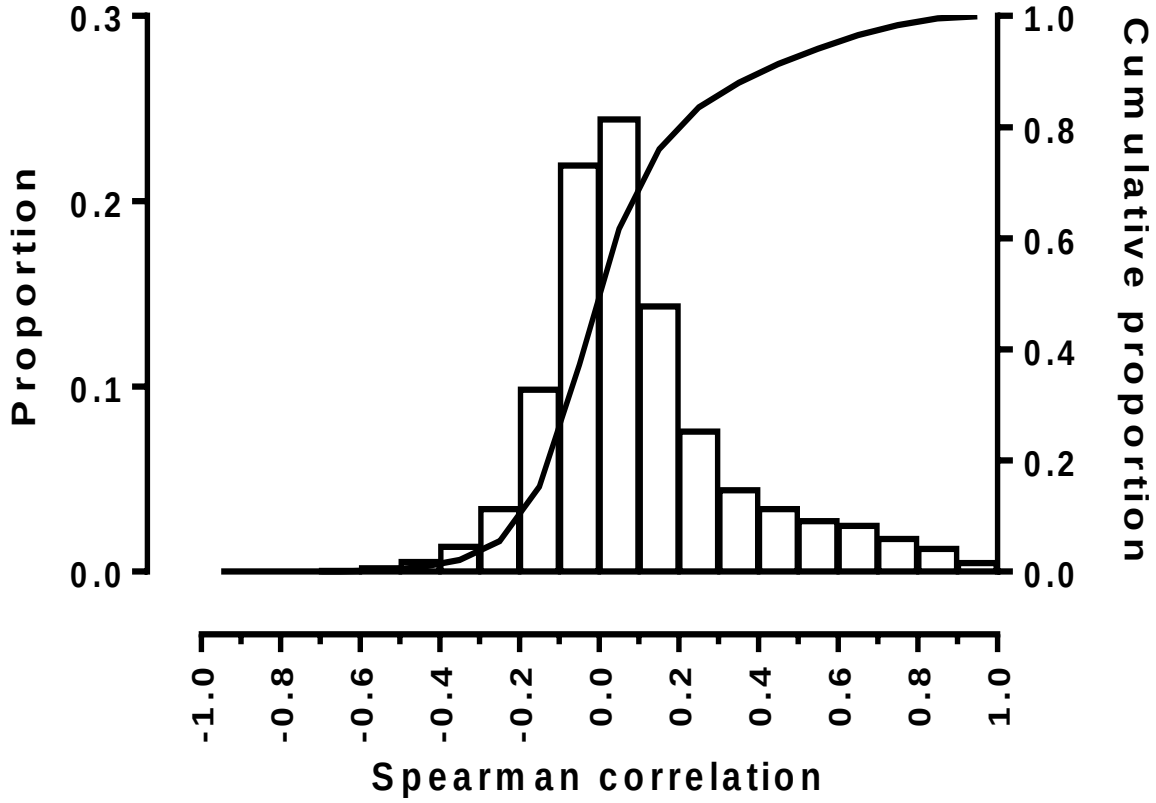
**David Torgerson (University of York, UK)**

**Mari Imamura, Research Fellow and native Japanese speaker (University of Aberdeen, UK)**

**Doug Altman and anonymous referees for their comments**

**Numerous colleagues who listened and advised**

**Figure 1: Distribution of pairwise correlations between baseline values in 13 genuine randomised controlled trials.**




**The bars show the proportion and the line the cumulative proportion.**

**Figure 3: Results from 1000 simulated randomisations in 13 genuine randomised controlled trials.**

**The top panel shows the distribution of p-values from comparisons of baseline variables (1000 simulations of 726 variables). The dotted line is the expected proportion of 0.10. The error bars are the 95% confidence interval for each decile. The shaded area shows the "normal" range (7.4%-12.7%) that contains 95% of the proportions for individual deciles in the 1000 simulations. The middle panel shows the percentage of simulations in which the proportion of p-values for each decile lay outside this "normal" range. The bottom panel shows the number of deciles with proportions outside the "normal" range in a single simulated randomisation for the 1000 simulations.**

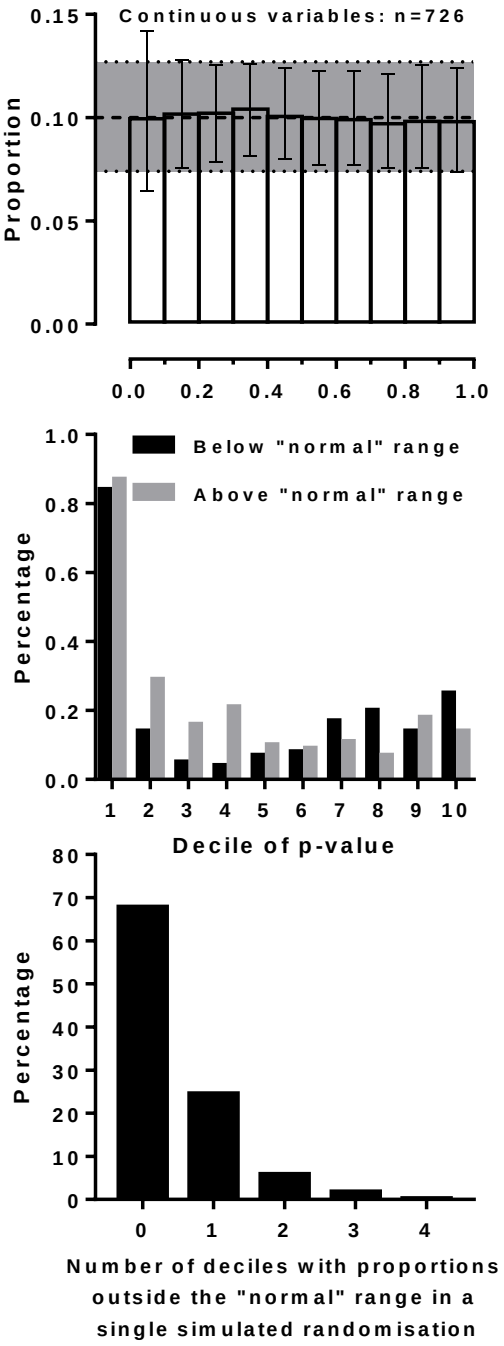


**Figure 4: Distribution of baseline p-values from categorical variables in 13 genuine randomised controlled trials.**

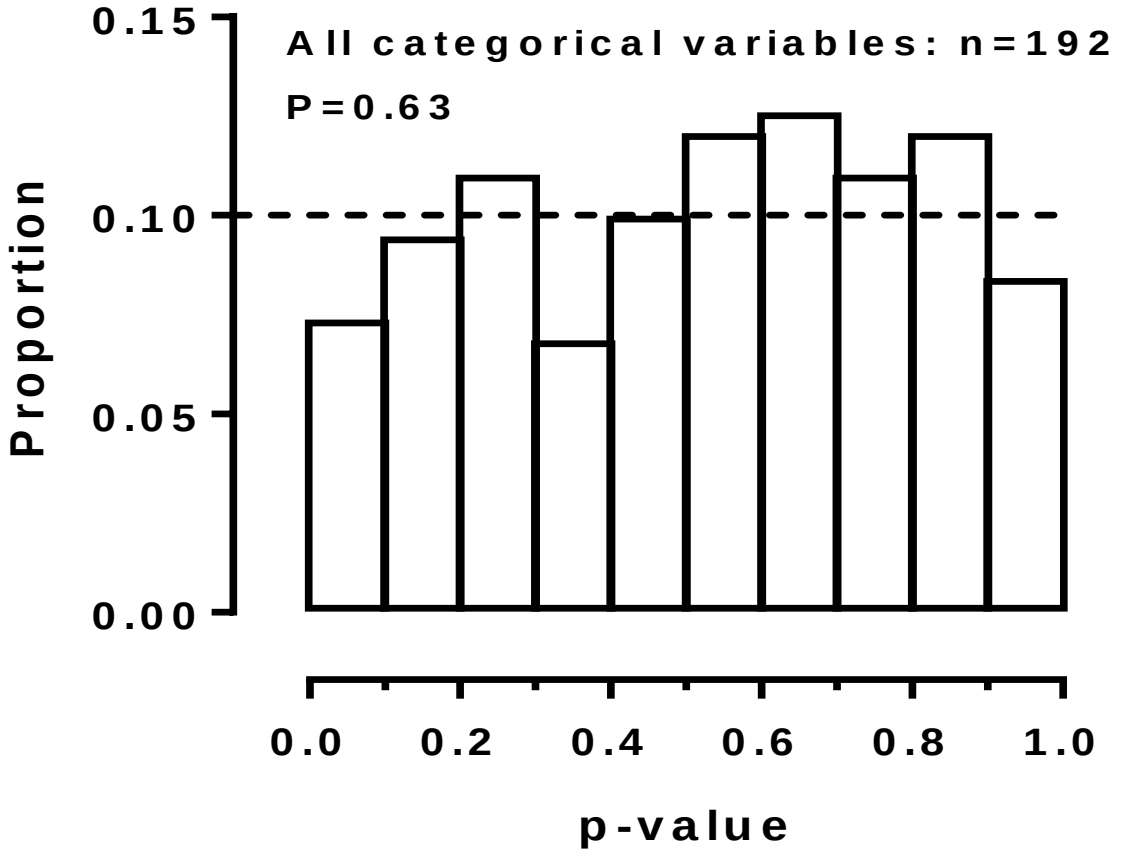


**Figure 5: Distribution of baseline p-values from all variables in 13 genuine randomised controlled trials.**

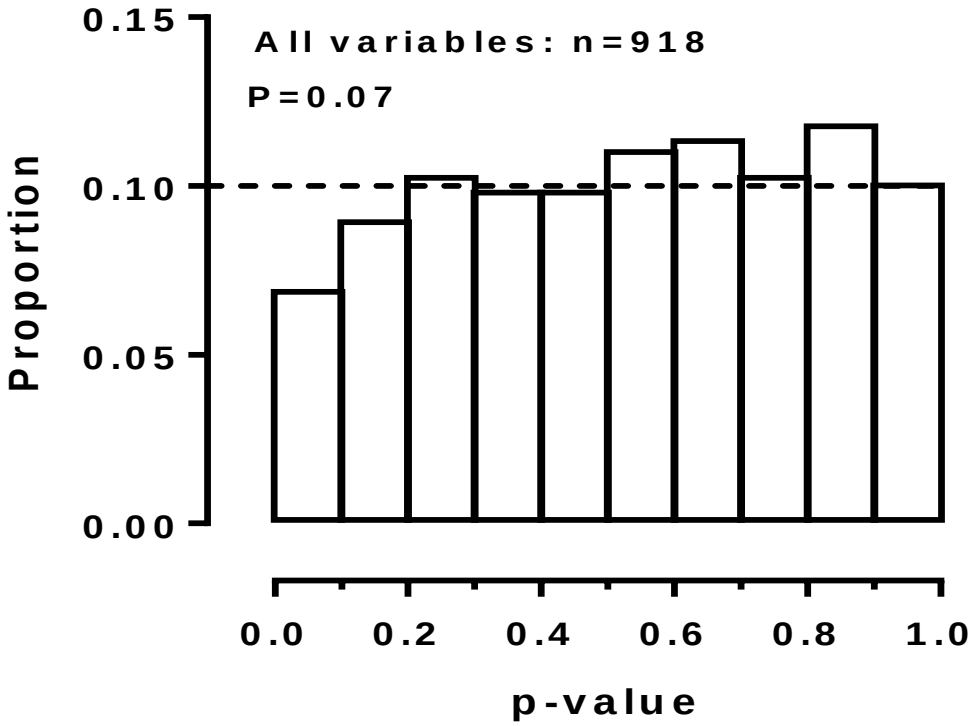
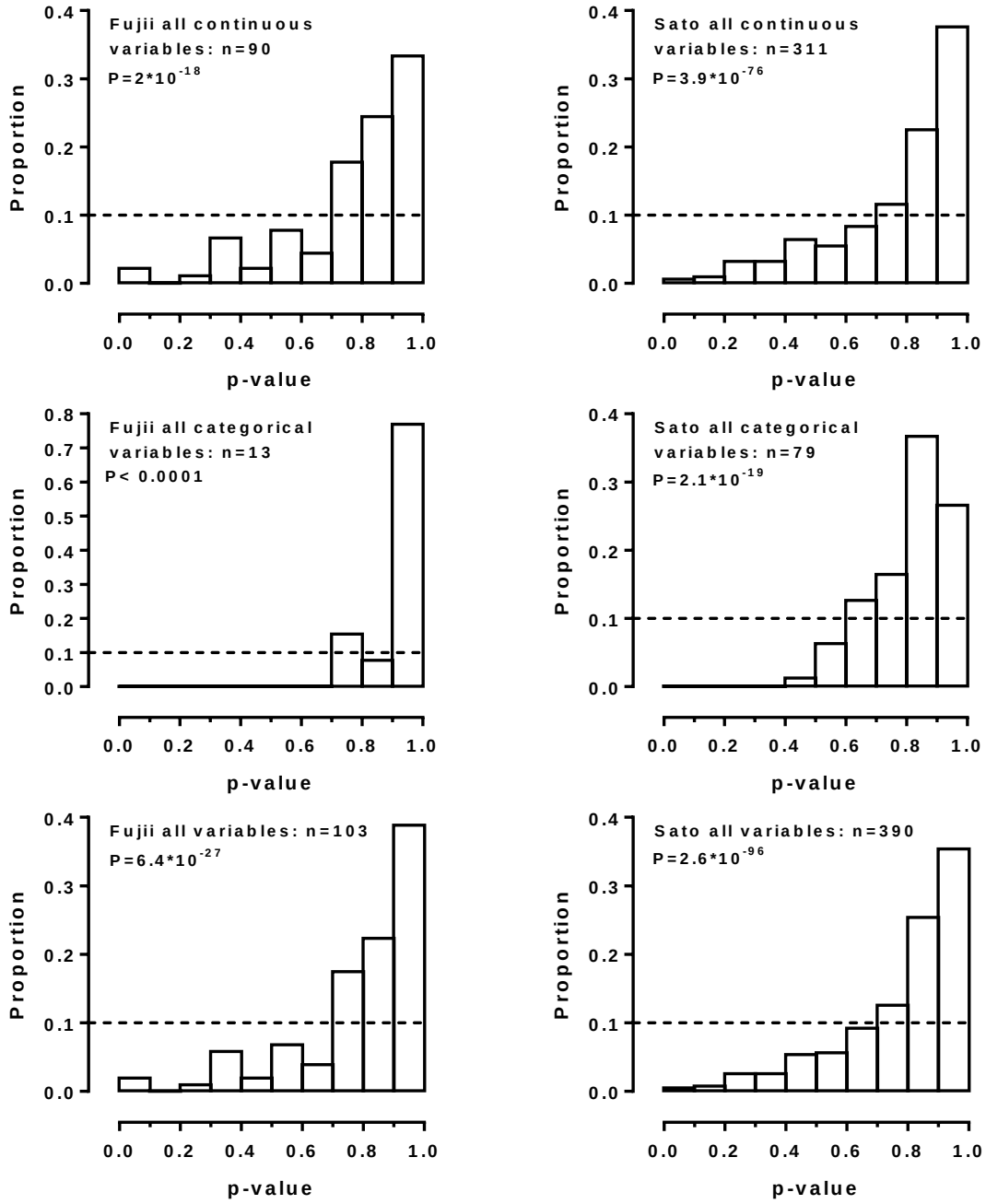


**Figure 6: Distribution of baseline p-values in randomised controlled trials by Fujii and Sato.**



## Figure 7: P-values for the distribution of baseline p- values from genuine randomised controlled trials and trials containing fabricated data

**The top panel shows the p-values from the one-way chi square tests for the distribution of baseline p-values for continuous variables by decile for individual randomised controlled trials with at least 10 variables by Sato, Fujii and 13 genuine randomised trials from our group in Auckland. The bottom panel shows the p-values for the distribution of baseline continuous variables available for at least 10 studies by decile. The open squares represent retracted trials. The Auckland trial analyses are restricted to 30 commonly reported variables in the Box.**