Incorrectly analysing stratified and minimised trials may lead to wrongfully rejecting superiority of interventions

It is with great interest that we read the report of Yoshida *et al*[1] on the effect of second-generation narrow band imaging compared to white light imaging on detecting early gastric cancer in high-risk patients. The trial was expertly designed with a large patient population and, although superiority of narrow band imaging could not be proven, has important implications for further research on this topic. However, a significant issue concerning the analyses attracted our interest and we would like to comment on it.

The primary outcome, the difference in proportion of patients in whom early gastric cancer was diagnosed, failed to reach statistical significance (p=0.412). This difference in proportions was tested for significance using Fisher's exact test. This might not have been the proper method for analysis, as patients in the study were randomised using minimisation with a random component, stratified by institution, age and indication of endoscopy.

Imbalance of risk factors between treatment and control arms can occur by chance under normal randomisation, possibly leading to confounded treatment estimates. Stratification and minimisation are useful methods to ensure balance of risk factors between treatment arms.[2–4] These methods can be beneficial in small and large trials, but for trials larger than 1000 patients little effect of minimisation on imbalance was found as compared with simple randomisation.[4]

One of the assumptions of Fisher's exact test is that samples are random and independent, which is not the case in this study. The problem that occurs with stratified or minimised randomisation is clustering between treatment groups which introduces positive correlation between observations. The correlation between observations violates the independence assumption and will lead to standard errors (SE) that are biased upwards because tests for independent samples do not account for this correlation and will overestimate the variance of the treatment effect.

As a SE that is biased upwards leads to inflated p values, not accounting for these balancing variables in the analysis may lead to wrongfully not rejecting the null hypothesis. This effect can be considerable, as a reanalysis of a large trial showed twofold to fourfold increases in p values and a simulation study showed reductions in power of up to 30 percentage points.[5 6] Thus, studies using either of the balancing methods should adjust their analysis for the balancing variables used in the randomisation procedure.[5]

Considering the report of Yoshida *et al*, this could mean that they may have erroneously concluded that second-generation narrow band imaging was not superior to white light imaging. We cannot determine the precise effect adjustment would have had in the study by Yoshida *et al* as a reanalysis requires the individual patient data. While adjustment is possible using Fisher's exact test,[7 8] we suggest performing logistic regression analysis as this allows adjustment for multiple minimisation variables, does not rely on inefficient stratification,[8] and can be used to determine the confidence interval around the treatment effect estimate.

Unadjusted analyses in balanced randomised trials seem to be a recurring phenomenon. In 2012, a systematic review showed that only 26% of trials published in leading journals that used a balancing method correctly adjusted for all balancing factors.[6] But even in more recent trials analyses are often not adjusted for balancing factors, as shown by the study of Yoshida *et al*, but also by other trials in the leading journals of gastroenterology and hepatology.[1 9 10] When used correctly minimisation and stratification are powerful tools for balancing randomised trials and improving the validity of studies. However, this has important consequences for data analysis. As such, we urge trialists to include the balancing variables as adjustment factors in their statistical analyses.

Reinier Cornelis Anthonius van Linschoten ,[1,2] **Rachel West**,[1] **Desirée van Noord**,[1] **Nikki van Leeuwen**[3]

[1]Gastroenterology & Hepatology, Franciscus Gasthuis en Vlietland, Rotterdam, The Netherlands
[2]Gastroenterology & Hepatology, Erasmus Medical Center, Rotterdam, The Netherlands
[3]Public Health, Erasmus Medical Center, Rotterdam, The Netherlands

Correspondence to Reinier Cornelis Anthonius van Linschoten, Gastroenterology & Hepatology, Franciscus Gasthuis en Vlietland, Rotterdam, The Netherlands; r.linschoten@franciscus.nl

Contributors RCAvL drafted the manuscript. All authors critically reviewed the manuscript and approved the final version of the manuscript for submission.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests RCAvL has nothing to disclose. DvN reports grants from AbbVie, Falk, Ferring, Janssen, MSD, Pfizer, and Takeda and personal fees from Janssen and Takeda outside the submitted work. RW reports grants from AbbVie, Falk, Ferring, Janssen, MSD, Pfizer, and Takeda and personal fees from AbbVie, Janssen and Pfizer outside the submitted work. NvL Leeuwen has nothing to disclose.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; internally peer reviewed.

To cite van Linschoten RCA, West R, van Noord D, *et al*. *Gut* 2022;**71**:1038–1039.

Received 14 April 2021
Accepted 23 June 2021
Published Online First 14 July 2021

Gut 2022;**71**:1038–1039. doi:10.1136/gutjnl-2021-324936

ORCID iD
Reinier Cornelis Anthonius van Linschoten http://orcid.org/0000-0003-3052-596X

REFERENCES

1 Yoshida N, Doyama H, Yano T, *et al*. Early gastric cancer detection in high-risk patients: a multicentre randomised controlled trial on the effect of second-generation narrow band imaging. *Gut* 2021;70:67–75.
2 Treasure T, MacRae KD. Minimisation: the platinum standard for trials? *BMJ* 1998;317:362–3.
3 Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ* 2005;330:843.
4 Scott NW, McPherson GC, Ramsay CR, *et al*. The method of minimization for allocation to clinical trials. A review. *Control Clin Trials* 2002;23:662–74.
5 Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med* 2012;31:328–40.
6 Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. *BMJ* 2012;345:e5840.
7 Kalsih LA, Riester KA, Pocock SJ. Accounting for Non-Independent observations in 2×2 tables, with application to correcting for family clustering in Exposure-Risk relationship studies. *J Mod App Stat Meth* 2002;1:379–86.
8 Jung S-H. Stratified Fisher's exact test and its sample size calculation. *Biom J* 2014;56:129–40.
9 Cox SR, Lindsay JO, Fromentin S, *et al*. Effects of low FODMAP diet on symptoms, fecal microbiome, and markers of inflammation in patients with quiescent inflammatory bowel disease in a randomized trial. *Gastroenterology* 2020;158:176–88.
10 Teoh AYB, Kitano M, Itoi T, *et al*. Endosonography-guided gallbladder drainage versus percutaneous cholecystostomy in very high-risk surgical patients with acute cholecystitis: an international randomised multicentre controlled superiority trial (DRAC 1). *Gut* 2020;69:1085–91.