Guidelines for Reporting Study Results

Prof. Tony Panzarella CHL5207Y/08Y
Jan. 12, 2017

Example: Walker et al. Lancet Oncology 2015

Pretreatment with anti-thymocyte globulin versus no anti-thymocyte globulin in patients with haematological malignancies undergoing haemopoietic cell transplantation from unrelated donors: a randomised, controlled, open-label, phase 3, multicentre trial

Irwin Walker, Tony Panzarella, Stephen Couban, Felix Couture, Gerald Devins, Mohamed Elemary, Geneviève Gallagher, Holly Kerr, John Kuruvilla, Stephanie J Lee, John Moore, Thomas Nevill, Gizelle Popradi, Jean Roy, Kirk R Schultz, David Szwajcer, Cynthia Toze, Ronan Foley, on behalf of the Canadian Blood and Marrow Transplant Group

Abstract

Summary

Background Pretreatment with anti-thymocyte globulin (ATG) decreases the occurrence of chronic graft-versus-host disease (CGVHD) after haemopoietic cell transplantation from an unrelated donor, but evidence of patient benefit is absent. We did a study to test whether ATG provides patient benefit, particularly in reducing the need for long-term immunosuppressive treatment after transplantation.

Methods We did a phase 3, multicentre, open-label, randomised controlled trial at ten transplant centres in Canada and one in Australia. Eligible patients were aged 16 to 70 years with any haematological malignancy and a Karnofsky score of at least 60 receiving either myeloablative or non-myeloablative (or reduced intensity) conditioning preparative regimens before haemopoietic cell transplantation from an unrelated donor. We allocated patients first by simple randomisation (1:1), then by a minimisation method, to either pretransplantation rabbit ATG plus standard GVHD prophylaxis (ATG group) or standard GVHD prophylaxis alone (no ATG group). We gave a total dose of ATG of 4·5 mg/kg intravenously over 3 days (0·5 mg/kg 2 days before transplantation, 2·0 mg/kg 1 day before, and 2·0 mg/kg 1 day after). The primary endpoint was freedom from all systemic immunosuppressive drugs without resumption up to 12 months after transplantation. Analysis was based on a modified intention-to-treat method. This trial was registered at ISRCTN, number 29899028.

Findings Between June 9, 2010, and July 8, 2013, we recruited and assigned 203 eligible patients to treatment (101 to ATG and 102 to no ATG). 37 (37%) of 99 patients who received ATG were free from immunosuppressive treatment at 12 months compared with 16 (16%) of 97 who received no ATG (adjusted odds ratio 4·25 [95% CI 1·87–9·67]; p=0·00060. The occurrence of serious adverse events (Common Terminology Criteria grades 4 or 5) did not differ between the treatment groups (34 [34%] of 99 patients in the ATG group *vs* 41 [42%] of 97 in the no ATG group). Epstein-Barr virus reactivation was substantially more common in patients who received ATG (20 [one of whom died—the only death due to an adverse event]) versus those who did not receive ATG (two [no deaths]). No deaths were attributable to ATG.

Interpretation ATG should be added to myeloblative and non-myeloblative preparative regimens for haemopoietic cell transplantation when using unrelated donors. The benefits of decreases in steroid use are clinically significant. Epstein-Barr virus reactivation is increased, but is manageable by prospective monitoring and the use of rituximab. Future trials could determine whether the doses of ATG used in this trial are optimum, and could also provide additional evidence of a low relapse rate after non-myeloablative regimens.

Funding The Canadian Institutes of Health Research and Sanofi.

Introduction

- State the purpose of the study.
  - explanatory versus pragmatic study.

Pragmatic research

- Asks whether an intervention works under real-life conditions, and whether it works in terms that matter to the patient
- Concerned with whether the intervention works, not how or why

Explanatory research

- Asks whether an intervention works under ideal or selected conditions
- It is more concerned with how and why an intervention works
- Valuable for understanding questions of efficacy
- Limited value in determining whether to provide a service to a wide variety of patients in a wide variety of circumstances

Pragmatic vs. explanatory research

- Important distinction because it will determine key methodological issues
  - Patient selection
  - Definition of intervention and controls
  - Use of blinding and placebos
  - Choice of outcome
  - Type of analysis

Patient selection

- Pragmatic
  - Reflects Routine practice
  - Broad selection criteria (limited exclusion criteria)
  - Interested to know whether intervention works in general
- Explanatory
  - Recruitment more selective
  - Exclude patients with co-morbidities or a doubtful diagnosis
  - Can establish whether intervention works under ideal conditions

The intervention

- Pragmatic
  - Left much more to the discretion of the clinician
  - Provided by normal staff with routine training
- Explanatory
  - Strictly defined and tightly controlled
  - Provided by "experts"

The control

- Pragmatic
  - Not concerned with how an intervention works compared to a control; thus control can be routine practice
- Explanatory
  - Uses a placebo
  - Is there a placebo effect?

Blinding

- Desirable for an explanatory study
  - Avoids bias
  - If a patient is aware of their treatment group they may derive benefit from mechanisms other than the direct benefit of the treatment
    - E.g. Placebo effect
- May be inappropriate for a pragmatic trial

Outcome measures

- Pragmatic
  - Usually patient centred
  - Incorporate broad measures of quality of life (QOL)
  - Does the treatment work from the patients point of view?
- Explanatory
  - May focus on specific symptoms or dimensions of QOL
  - May include clinical or biologic measures
  - May not relate directly to patients preference

Type of analysis

- Pragmatic
  - Intention to treat
  - Measuring the effect of making the treatment available rather than actually receiving treatment
- Explanatory
  - As treated
  - Risk of bias

Pragmatic or explanatory?

- Represent ends of a continuum rather than distinct entities
  - E.g. pragmatic trial with broad selection criteria, patient-centred outcome measure and intention to treat analysis may also include explanatory elements - subgroup analysis

Efficacy vs. effectiveness

- Not the same thing
- Efficacy – does the intervention work under ideal conditions
  - Can this work?
- Effectiveness – does the intervention work under routine or "real life" conditions
  - Does this work?

Introduction

- If the study was designed to test one or more *a priori* hypotheses, state the hypotheses.

- Distinguish between the primary and secondary hypotheses.

Introduction

- Identify any planned subgroup or secondary analyses.

- Results of post-hoc (after-the-fact) analyses should be considered as exploratory or speculative.

Introduction

Our trial focuses for the first time on CGVHD and its effect on quality of life, and includes patients receiving both non-myeloablative and myeloablative conditioning.

Introduction

The trial tests the hypothesis that pretreatment with ATG before haemopoietic cell transplantation provides patient benefit in addition to decreasing the incidence of CGVHD.

Materials and Methods

- Indicate the unit of analysis.
  - usually the patient.
- Describe the population to be studied and to which the results are to be generalized.
- Provide definitions for all explanatory and response variables.
  - specify units of measurement.

Materials and Methods

- Specify the minimum change or difference in the response variable that is considered to be clinically important.

- Indicate that the study was approved by the appropriate REB.

Materials and Methods

- Describe the study design.
  - observational versus experimental.
  - retrospective versus prospective.
  - cross-sectional versus longitudinal.

- Describe the treatment under study and the protocol under which it was administered.

Materials and Methods

- Describe any potential confounding variables and the methods used to control for them.
  - can be controlled for by appropriate designs or statistical analysis.

- Identify the study setting and the source of study participants.

Materials and Methods

- Provide the rationale for the sample size calculation.
  - *every* study, even retrospective studies, should specify how the sample size was determined.

- Provide the inclusion and exclusion criteria for the study.
  - Inclusion criteria = explanatory variables of interest.
  - exclusion criteria = potential confounding variables to avoid.

Materials and Methods

- Describe the circumstances under which the informed consent was obtained.

- Specify how study participants were assigned to treatment groups.
  - for RCT's specify the method of random assignment.
  - how was allocation concealed?

Materials and Methods

- Specify the technique of masking (blinding), if applicable.
  - report whether masking was effective.
  - indicate who was blinded (single versus double versus triple blinding).

- Describe fully any placebo medications, or alternative or concomitant treatments received by control groups.

Materials and Methods

- Describe the methods of data collection or measurement.
- When data consist of observations/judgments identify:
  - training and experience of evaluator(s).
  - was the evaluator masked?
- For equipment used identify:
  - name, model number, calibration details, accuracy of measurements.

Materials and Methods

- Describe the planned nature and duration of follow-up.
  - is the follow-up period long enough to allow clinically meaningful observations?

- Describe any quality-control methods used to ensure completeness and accuracy of data collection.
  - e.g. double data entry

Statistical methods

- Describe the comparisons to be made and the statistical procedures to be used for making them.
  - Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results.
  - Preferably specified before the data are collected, but specificity not always possible.

Statistical methods

- Most statistical techniques have assumptions underlying their use. Indicate that these assumptions have been checked.
  - e.g. normality of distributions of variables analyzed with a *t*-test.

Statistical methods

- State whether the statistical analysis will be on the basis of intention-to-treat.
  - Results are analyzed based on random assignment.
  - A method of adjusting for bias caused by participants leaving the study *because* of the treatment under study.

Statistical methods

- Describe any planned interim analyses and any stopping rules.
  - Often an ethical imperative to stop the study if there is strong evidence that an experimental treatment is superior.
  - The release of interim results needs to be handled carefully.
  - Need to account for multiple testing.

Statistical methods

- Specify any procedures used to control for the multiple testing problem.
  - The more statistical analyses performed on the same data, the more likely *P* values will be wrongly accepted as indicating a biological relationship.
  - Need to adjust the alpha level, and report technique used (e.g. Bonferroni correction)

Statistical methods

- Report the levels of alpha (α) and beta (β) or statistical power, 1 - β).
  - A Type I error (alpha level) occurs when we reject the null hypothesis of no difference when in fact there is no difference.
  - A Type II error (beta level) occurs when we fail to reject the null hypothesis of no difference when in fact there is a difference of the specified size.

Statistical methods

- Report whether statistical tests were one- or two-tailed.
  - Two-tailed tests are more conservative requiring a larger treatment effect to achieve the same level of statistical significance.
  - The rationale for using a one-tailed test should be given.
  - Because the two types of tests produce different *P* values for the same data, the type of test must be specified a priori.

Statistical methods

- Identify the statistical package or program used to analyze the data.
  - Although commercial programs generally are validated and updated, privately developed programs may not be.
  - Not all statistical software uses the same algorithms to compute the same statistics.

Sample size

We calculated that a sample size of 196 assessable recipients would have a power of 0·8 to detect as significant a decrease in the use of immunosuppression from 60% in the no ATG group to 40% in the ATG group, based on a type I error (two-sided) of 0·05. We postulated that the addition of ATG to the preparative regimen would result in a 20% decrease in the proportion of recipients needing immunosuppression for CGVHD without an increase in mortality, disease relapse, or serious infection.

Study design

Study design and participants

We did a randomised, multicentre, open-label, phase 3 trial comparing pretransplantation rabbit ATG plus standard GVHD prophylaxis with standard GVHD prophylaxis alone, in recipients receiving either myeloablative or non-myeloablative conditioning before haemopoietic cell transplantation from unrelated donors. We completed the study in ten transplantation centres in Canada and one in Australia (appendix). This study was approved by Health Canada, the Australian Government, Department of Health Therapeutic Goods Administration, and Institutional Review Boards at each participating centre.

Primary endpoint

Outcomes

The primary efficacy endpoint was freedom from systemic immunosuppressive drugs without resumption up to 12 months after transplantation. This endpoint has

Secondary endpoints

Secondary endpoints included the time to neutrophil and platelet engraftment (analysed only for patients undergoing myeloablative conditioning); the incidence of AGVHD; the incidence of and time to CGVHD; times to non-relapse mortality, all-cause mortality, and relapse of haematological malignancy; event-free survival; graft rejection or failure (yes *vs* no); incidence of serious infection and cytomegalovirus reactivation; doses of immunosuppressive therapy at 12 months; and recipient-reported outcomes (quality of life).

Statistical analysis

Statistical analysis

We calculated that a sample size of 196 assessable recipients would have a power of 0·8 to detect as significant a decrease in the use of immunosuppression from 60% in the no ATG group to 40% in the ATG group, based on a type I error (two-sided) of 0·05. We postulated that the addition of ATG to the preparative regimen would result in a 20% decrease in the proportion of recipients needing immunosuppression for CGVHD without an increase in mortality, disease relapse, or serious infection.

Our analysis followed a modified intention-to-treat principle. Participants were excluded from analysis when a transplantation was not done, either because the patient withdrew consent, had a relapse, or died before the scheduled transplantation. We compared the primary endpoint between treatment groups using logistic regression adjusted for covariates.[24] Comparisons of time to failure endpoints incorporated Kaplan-Meier probability estimates, log-rank testing and, when adjustment of covariates was needed, use of the Cox proportional hazards model. The goodness-of-fit of the logistic regression model

Results

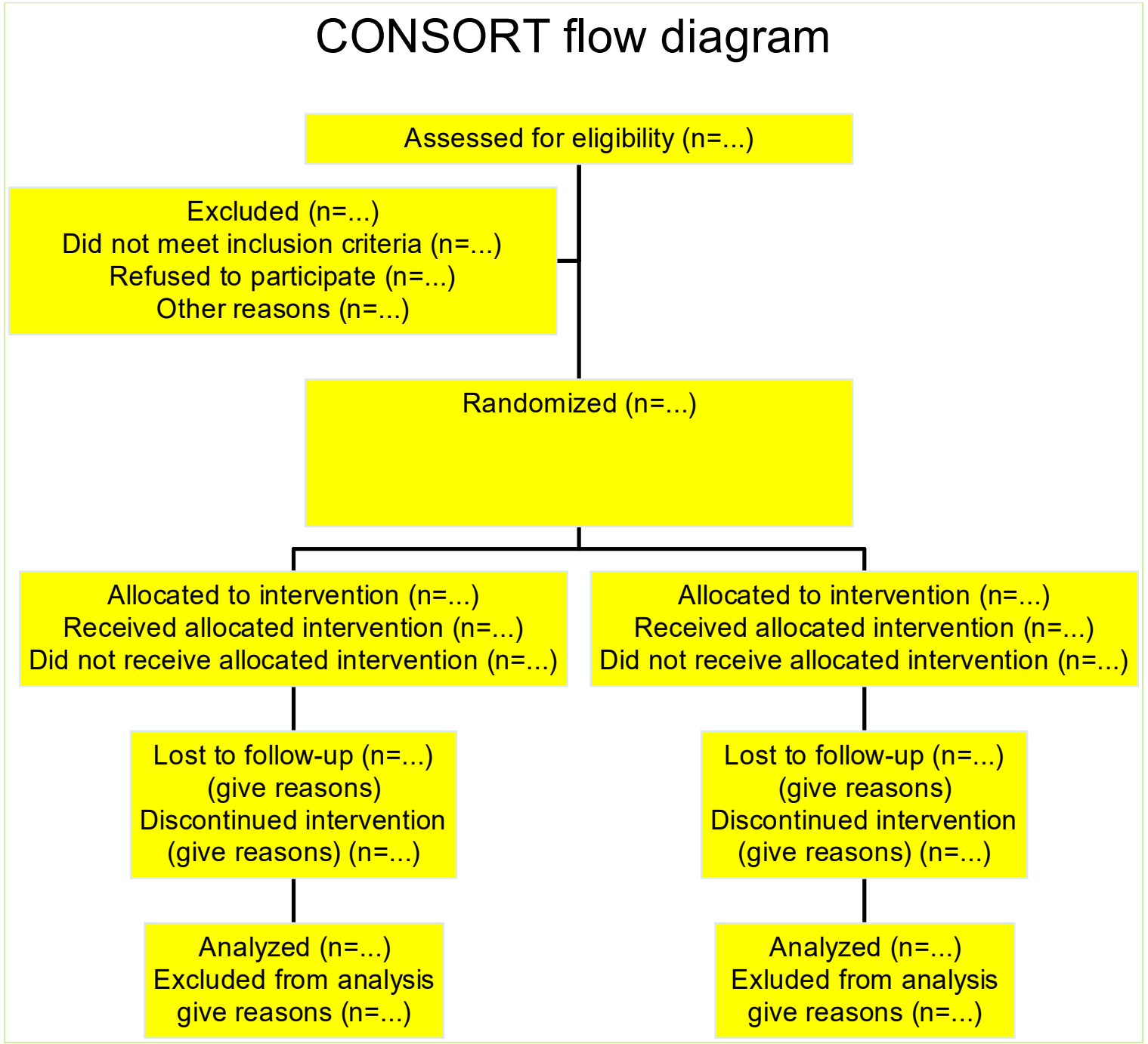
- Specify the beginning and ending dates of the data collection period and give the reasons for selecting those dates.
  - For retrospective studies identify the data accrual period, not the dates when the data were abstracted.

Results

- When possible provide a schematic summary of the study showing the number and disposition of participants at each stage.
  - # evaluated but who did not meet inclusion criteria.
  - # eligible but declined participation.
  - # on study who did not complete treatment.
  - # completed treatment but lost to follow-up.
  - # completed treatment and the follow-up period.

Consort Statement

- *CON*solidated *S*tandards *O*f *R*eporting *T*rials.
- Checklist for reporting randomized controlled trials (Moher *et al.*, *Annals of Internal Medicine*, 2001) .
- Synthesis of guidelines proposed by 2 working groups.
- See www.consort-statement.org


CONSORT flow diagram
Enrollment
Allocation
Follow-up
Analysis
Assessed for eligibility (n=...)
Excluded (n=...)
Did not meet inclusion criteria (n=...)
Refused to participate (n=...)
Other reasons (n=...)
Randomized (n=...)
Allocated to intervention (n=...)
Received allocated intervention (n=...)
Did not receive allocated intervention (n=...)
Allocated to intervention (n=...)
Received allocated intervention (n=...)
Did not receive allocated intervention (n=...)
Lost to follow-up (n=...)
(give reasons)
Discontinued intervention
(give reasons) (n=...)
Lost to follow-up (n=...)
(give reasons)
Discontinued intervention
(give reasons) (n=...)
Analyzed (n=...)
Excluded from analysis
give reasons (n=...)
Analyzed (n=...)
Exluded from analysis
give reasons (n=...)

Results

- Indicate the similarities and differences between the control group and the experimental group at baseline.
  - often listed as Table 1 in manuscripts
  - if baseline variables similar in the 2 groups then differences in response variable more easily attributable to the treatment under study

Results

- For observations based on judgments provide a measure of consistency or agreement among the evaluators
  - e.g. interpretation of radiographs by two or more evaluators.

Results

- Present the data and the analyses of the primary endpoint(s) first. Summarize the data for the primary comparison with descriptive statistics.
  - Satisfies readers' expectations
  - Helps to avoid claims of data dredging.

Results

- Second, report all other clinically relevant outcomes, whether expected or not.

- Third, describe any secondary or post hoc analyses that yielded interesting results.

Results

- Report absolute changes (and relative, if desirable) or differences for all primary endpoints.
  - Reporting relative changes could be misleading
  - A change from, say, 2 kg to 1 kg is the same 50% percent reduction as a change from 2000 kg to 1000 kg, but obviously they mean different things.

Results

- Report actual *P* values for all primary analyses.
  - Avoid $P < 0.05$. A $P = 0.051$ should be interpreted similarly as $P = 0.049$
  - Report *P* values to no more than two significant digits
  - A $P = 0.001$ is the smallest value that needs to be reported. If smaller use $P < 0.001$

Results

- Report (95%) confidence intervals for changes or differences in the primary endpoints.
  - As study results are estimates of the entire population of interest confidence intervals indicate the precision of the estimate.

Results

- Whenever possible, present main findings in figures or tables.
  - It saves space.
  - Allows for more information, and more clear information.
  - Avoid duplication of information in the text.

Results

- When feasible, report statistical findings with enough detail to allow subsequent re-analysis or meta-analysis.
  - For small studies report raw data.
  - Remember an article is a source of data for other researchers.

Results

- Indicate the degree to which study participants adhered to the protocol and explain any deviations from the protocol.
  - Deviations can produce bias.

- Report all potential treatment-related side effects and adverse events.

Results

- Describe how extreme values were handled.
  - Add variability and uncertainty to results.
  - Consider analyzing with and without these observations.

- Account for all observations and explain missing data.

Flow Diagram

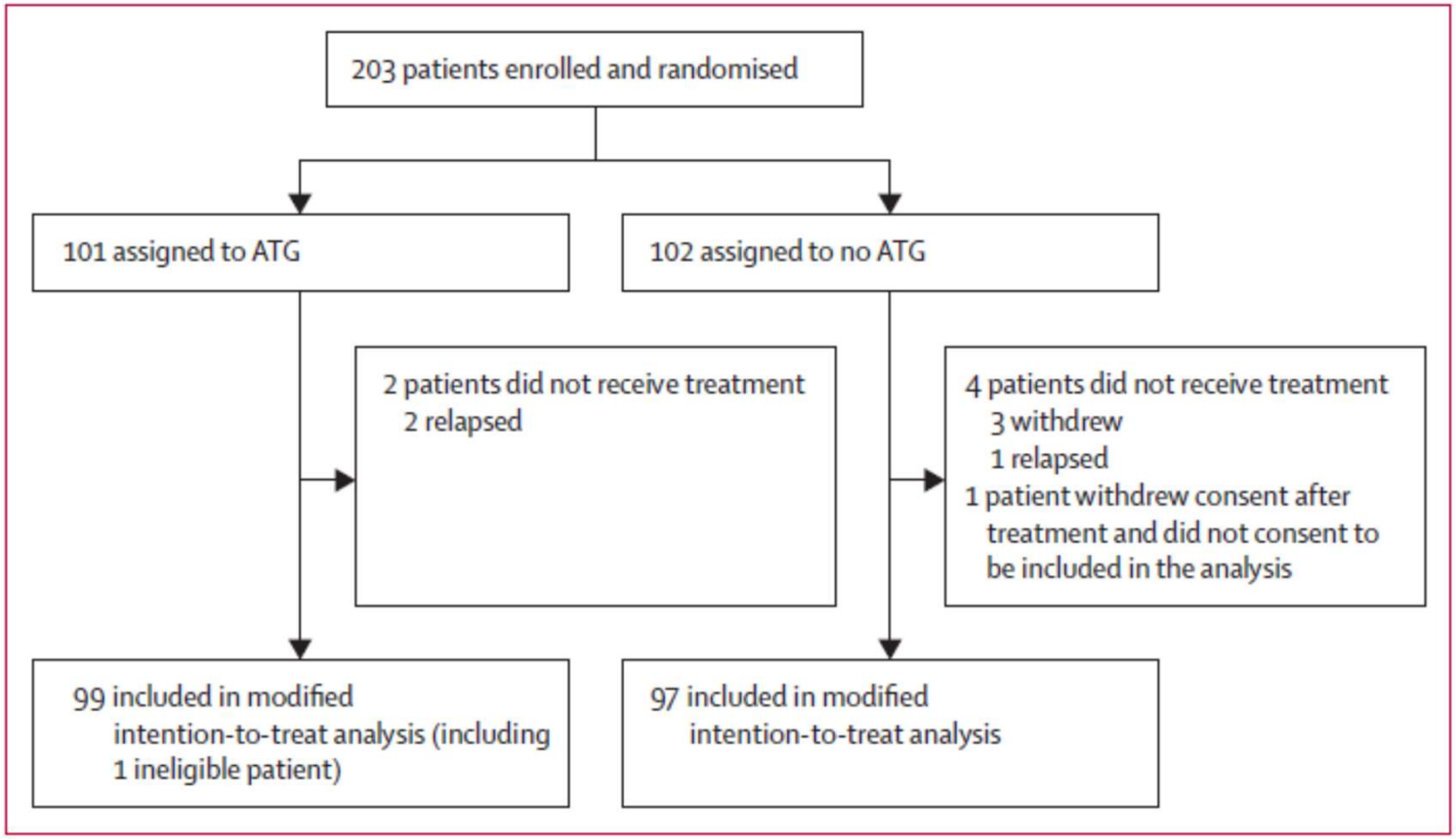


Figure 1: **Trial profile**
ATG=anti-thymocyte globulin.

Table 1

	No ATG (n=97)*	ATG (n=99)
Haematological disease		
Chronic myeloid leukaemia	5 (5%)	7 (7%)
Acute myeloid leukaemia	39 (40%)	39 (39%)
Acute lymphocytic leukaemia	16 (17%)	13 (13%)
Myelodysplastic syndrome	12 (12%)	11 (11%)
Chronic lymphocytic leukaemia	8 (8%)	8 (8%)
Lymphoma	13 (13%)	14 (14%)
Other	4 (4%)	7 (7%)
Disease stage		
Early	59 (61%)	57 (58%)
Late	34 (35%)	34 (34%)
Other	4 (4%)	8 (8%)
Preparative regimen		
Myeloablative	66 (68%)	66 (67%)
Non-myeloablative or reduced intensity conditioning	31 (32%)	33 (33%)
Recipient age (years)		
Median (IQR)	49 (40–56)	49 (40–57)
16–30	10 (10%)	10 (10%)
31–50	44 (45%)	44 (45%)
>50	43 (44%)	45 (45%)
Donor age (years)		
Median (IQR)	31 (25–39)	30 (24–39)
<30	38 (39%)	43 (44%)
≥30	50 (52%)	43 (44%)
Missing	9 (9%)	13 (13%)
HLA match		
Full match	79 (81%)	83 (84%)
One antigen or allele mismatch	18 (19%)	16 (16%)
Recipient sex		
Male	65 (67%)	63 (64%)
Female	32 (33%)	36 (36%)
Karnofsky performance score		
<80	2 (2%)	6 (6%)
80–90	58 (60%)	50 (51%)
100	37 (38%)	43 (43%)
Donor–recipient sex match		
Female donor–male recipient	16 (16%)	16 (16%)
Other	81 (84%)	83 (84%)
Comorbidities (HCT-CI)		
0	50 (52%)	61 (62%)
1–2	25 (26%)	15 (15%)
≥3	22 (23%)	23 (23%)

(Table 1 continues in next column)

Subgroup analysis

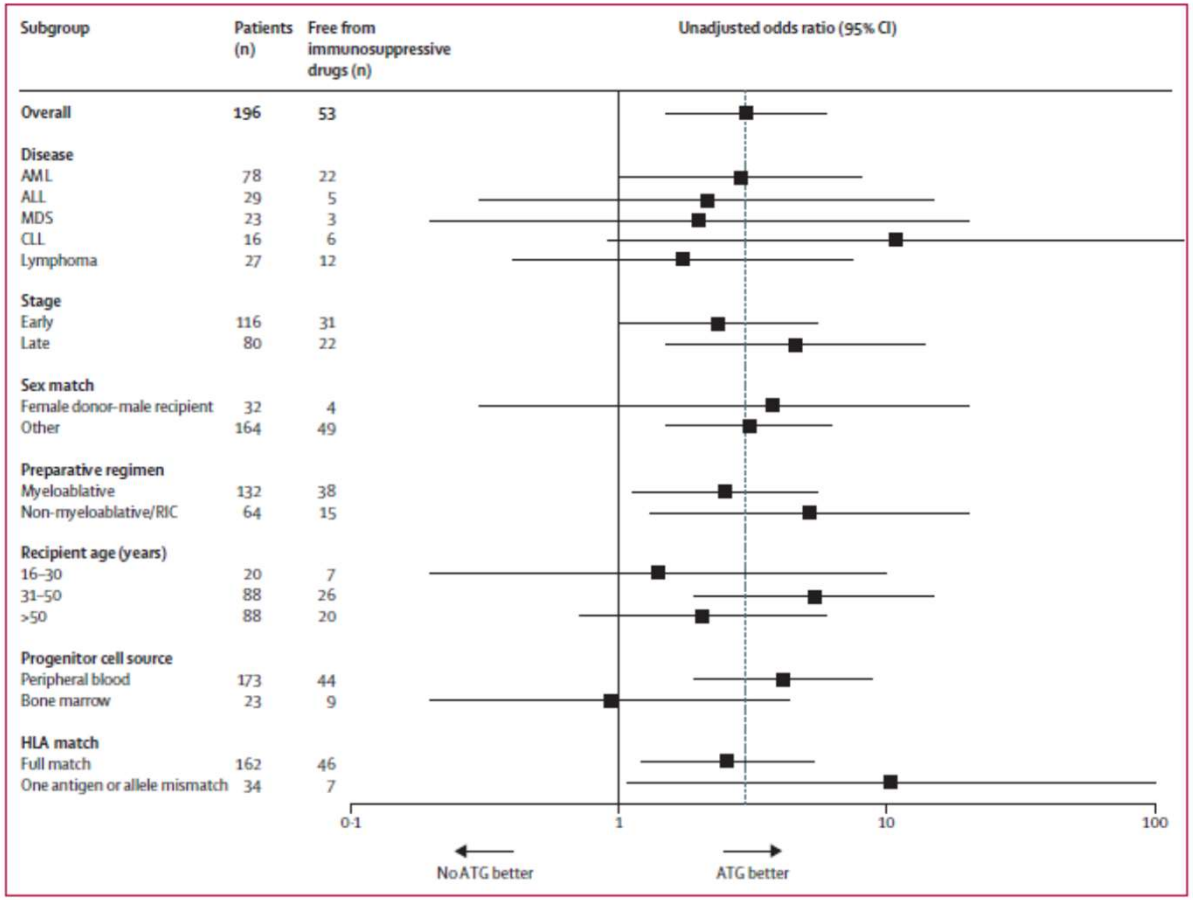


Figure 2: **Subgroup analyses for withdrawal of all systemic immunosuppressive drugs without resumption up to 12 months after transplantation**
AML=acute myeloid leukaemia. ALL=acute lymphocytic leukaemia. MDS=myelodysplastic syndrome. CLL=chronic lymphocytic leukaemia. Non-myeloablative/RIC=non-myeloablative or reduced intensity conditioning. ATG=anti-thymocyte globulin.

Figures

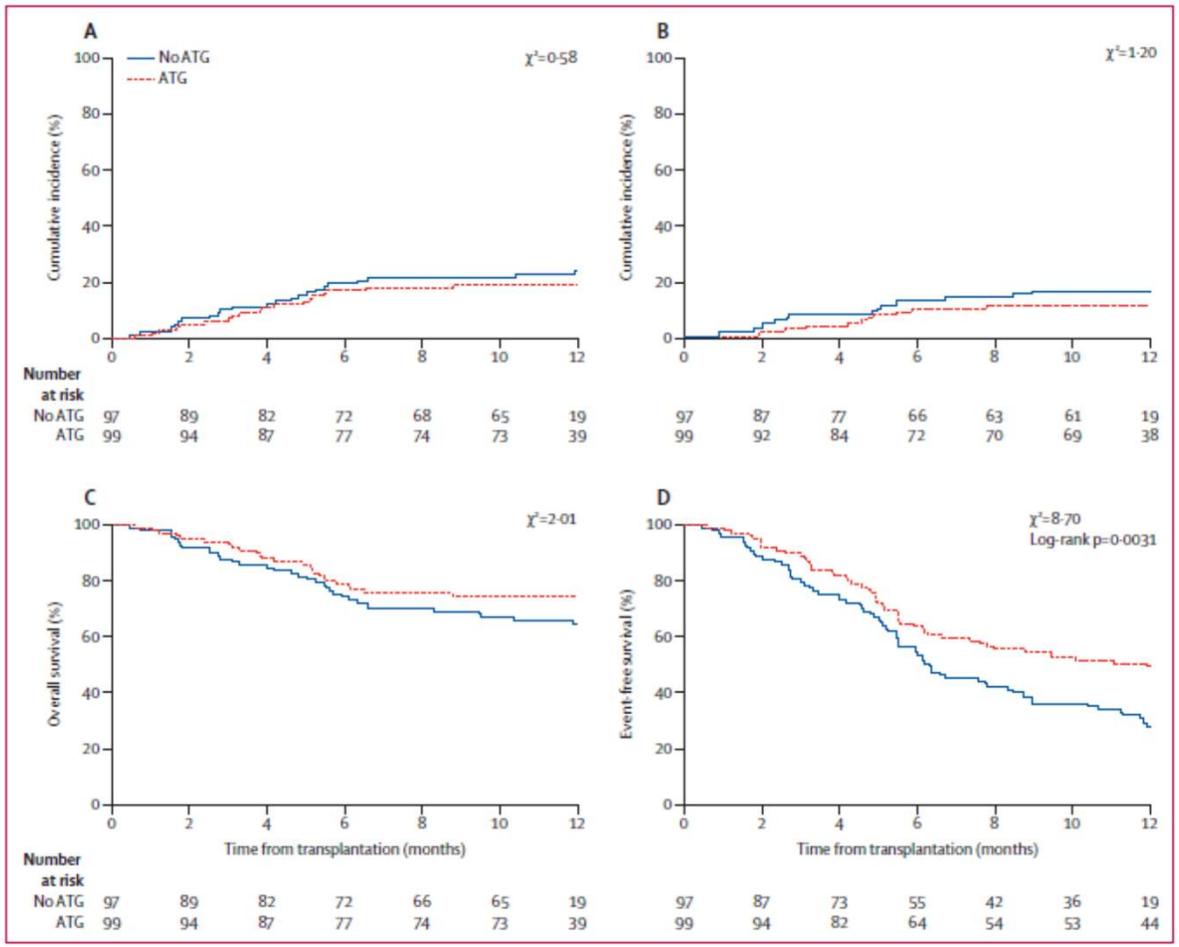


Figure 3: **Secondary efficacy endpoints**
(A) Incidence of non-relapse mortality. (B); Incidence of relapse. (C) Overall survival. (D) Event-free survival. ATG=anti-thymocyte globulin.

Secondary outcome: QoL

	No ATG (n=63)		ATG (n=74)		p value
	Adjusted mean (SE)	n	Adjusted mean (SE)	n	
Atkinson Life Happiness Rating	6·16 (0·41)	62	7·40 (0·39)	70	0·027
Bradburn Affect Balance Scale	7·02 (0·25)	48	7·00 (0·22)	64	0·960
Positive affect	3·03 (0·20)	48	3·11 (0·17)	64	0·760
Negative affect	1·02 (0·14)	48	1·11 (0·13)	64	0·610
Center for Epidemiologic Studies Depression Scale	14·57 (1·19)	47	11·77 (1·02)	64	0·087
Illness Intrusiveness Ratings Scale					
Total score	45·19 (2·25)	47	41·73 (1·93)	64	0·250
Instrumental	4·45 (0·22)	47	4·31 (0·19)	65	0·630
Intimacy	3·60 (0·27)	45	3·60 (0·23)	61	0·990
Relationships and personal development	2·82 (0·18)	47	2·43 (0·15)	64	0·096
FACT-BMT	104·23 (2·64)	45	107·45 (2·21)	64	0·350
Trial outcome index	63·67 (1·98)	44	66·94 (1·64)	64	0·210
EuroQol-5D total score	0·78 (0·03)	45	0·82 (0·02)	62	0·250
CGVHD symptoms (Lee)					
Day 100	14·71 (1·39)	62	13·71 (1·19)	73	0·580
6 months	15·63 (1·73)	56	11·38 (1·28)	61	0·048
12 months	20·93 (2·19)	46	14·95 (1·35)	60	0·017

ATG=anti-thymocyte globulin. CGVHD=chronic graft-versus-host disease. FACT-BMT=Functional Assessment of Cancer Therapy scale and its Bone Marrow Transplant subscale.

Table 2: **Quality-of-life scores at 12 months after transplantation**

Discussion

- Discuss the implications of the primary analyses first.
- Distinguish between clinical importance and statistical significance.
- Discuss results in context of published literature.
- Discuss generalizability of results.

Discussion

- Discuss any weaknesses in the research design or problems with data collection, analysis, or interpretation.
  - Identifying difficult areas may help other researchers to avoid similar problems.

Discussion

Discussion

The results of this trial closely support our primary hypothesis that ATG prophylaxis would decrease the number of patients needing immunosuppressive treatment at 1 year after haemopoietic cell transplantation from unrelated donors. This effect was noted after both myeloablative and non-myeloablative conditioning in our study. The decrease in use of immunosuppressive therapy supports prospectively the secondary results from a long-term follow-up of a previously reported randomised trial.[4,6] In the long-term follow-up,[6] the probability of being alive and free of immunosuppressive drugs at 3 years was 53% in patients taking ATG and 17% in those not taking ATG. This trial differed from ours in that their main objective was to assess the incidence of AGVHD or death within 100 days of transplantation and that the investigators used a different ATG product (anti-Jurkat ATG).

Limitations

The limitations of this trial include the non-blinded design and short follow-up. Blinding was not feasible because of the frequent and obvious infusional reactions that occur with ATG,[36,37] but previous randomised trials of ATG have also used non-blinded designs.[4,5] A longer follow-up of this trial might reveal additional symptom burden because organ damage and immunosuppressive side-effects are expected to accumulate in patients affected by CGVHD.

Conclusions

- Should be limited to those supported by results.
- Results of secondary or post hoc analyses should be presented as exploratory.
- List conclusions and describe implications.
  - Allows readers to find them more easily.
  - Discussing implications puts study in perspective.

Conclusions

In summary, the results of this study strongly support the use of rabbit ATG at a dose of 4·5 mg/kg for unrelated donor haemopoietic cell transplantation by showing a markedly decreased incidence of immunosuppressive use, a decrease in patient-reported symptoms of CGVHD, and no difference in serious adverse events compared with no ATG use. Epstein-Barr virus reactivation is common, and monitoring patients together with either pre-emptive or prophylactic treatment with rituximab is advised.

References

- Lang TA and Secic M. How to Report Statistics in Medicine. American College of Physicians, Philadelphia, PA. 1997.

- Moher D, Schultz KF and Altman DG. The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *Ann Intern Med.* 2001;134:657-662.

- Brown J and Machin D. Statistics and Clinical Oncology. *Clinical Oncology* 1996;8:71-74.

- Walker I, Panzarella T, Couban S et al. Pretreatment with anti-thymocyte globulin versus no anti-thymocyte globulin inpatients with haematological malignancies undergoing haemopoietic cell transplantation from unrelated donors: a randomised, controlled, open-label, phase 3, multicentre trial. *Lancet Oncology* Dec 2015.