COSECSA Research Methodology Course Handbook

Jacob S Dreyer (Editor)

for

The College of Surgeons of East, Central and Southern Africa (COSECSA)

& The International Federation of Surgical Colleges (IFSC)

Alba CCCD

Published by Alba Critical Care Course Design, a Scottish Incorporated Charitable Organisation (Scottish Charity no SC043790), also known as Alba CCCD, 8 Woodgrove Avenue, Dumfries, DG1 1RR, UK. Alba CCCD is a registered publisher in the UK.

www.albacccd.com

albacccd@gmail.com

ISBN 978-0-9927099-6-9

Printed by Alba Printers, 1 St. Michael Street, Dumfries, DG1 2QD.

For Dr Brian Clark
Clinical Oncologist
"The Beatson"
Glasgow, UK
"for transforming hope into reality"

This book should be read with the PowerPoint slideshows available at www.albacccd.com or the COSECSA School for Surgeons site.

A.1: COSECSA Research Methodology Course (RMC): Online course Syllabus 2020 & Contents

Introduction

The following sections consist of PowerPoint slides of mini-lectures on topics explained in the detailed curriculum (section A.2), followed by step-by step practical sessions to help participants develop their own research ideas into a viable research question and proposal for a project.

The lectures should be viewed in the PowerPoint presentations that are available online; the slides in the handbook are really only there as quick reminder of lecture contents for those who have already worked through the presentations slide-by-slide. Presentations can be viewed individually so that each participant can learn at their own pace, or in small groups with discussion of each slide so that contentious points are clarified through discussion.

Practical sessions work best if done in small groups. The topics follow the just-delivered lectures in content and planning. The idea is that peers plan a study and develop a proposal together that they can present to their departmental head and/or the research committee at their institution, and therefore create their own viable project based on the principles learned during the course.

Quizzes and a final course test will be developed and made available **after the Covid-19 pandemic** has settled. If possible, the plan is then to have all presentations and quizzes written into an interactive online portal. Final course assessment will be through an online test for which participants will have to register at a fee decided by COSECSA, and completion will generate an automatic course certificate. Unfortunately it takes time to create the interactive online portal and the team members who are doing this are now all occupied due to the Covid-19 epidemic in UK.

A.2: Introduction to the COSECSA online Research Methodology Course (RMC)

Background

Sub-Saharan Africa is rich in surgical research material. Health care in the region deals with the interface between the effects of infectious diseases and extreme poverty, and the rise of non-communicable diseases. The WHO projection is that 80% of future cancer deaths will occur in low- and middle-income countries. At the same time the epidemics of serious and neglected injuries due to road-traffic crashes and burns need to be addressed, as well as the burden of congenital birth defects such as cleft lip and palate, club foot and reversible gastro-intestinal defects. In areas such as pre-hospital care, where simple advances can have major impact on survival from surgical emergencies, research based evidence can drive such advances.

After acceptance of resolution 68.15 by the World Health Assembly emergency and essential surgery has slowly become part of the global health agenda. Providing surgical care to everyone worldwide at acceptable levels is an almost impossible task unless evidence is produced to show where government expenditure will bring most benefit.

Quality improvement in clinical care and in patient safety in surgery happen more readily when surgeons produce their own evidence of improvements in care pathways, rather than being told by outsiders how they should change practice.

Surgeons in the COSECSA region demand to own more of their own intellectual property. A simple internet search for "funding opportunities for surgical research in Africa" showed that funding is easily available for students or trainees from rich countries to go to Africa on a short visit, collect some data for an audit study or clinical series and write it up as "research" that promote their own career, rather than helping care in the host country. Such studies usually produce no more than Level 3 evidence, and local collaborators are almost never first authors of such publications and then have to pay exorbitant fees to access their own research online.

A request therefore came in 2015 from COSECSA's Chair of the Education and Research Committee to the International Federation of Surgical Colleges (IFSC) to help develop a basic course in surgical research methodology for COSECSA trainees as no such course was then offered through COSECSA, although some trainees in the region attend such courses if they are also in an MMed programme. In 2015-16 the framework for such a course was written and pilot courses delivered in Ethiopia (twice), Rwanda and Zambia, with training of research methodology tutors. Unfortunately it is almost impossible to sustain another residential course due

to the costs of travel, accommodation and time out of work for trainees and tutors. In 2019 Dr Laston Chikoya therefore approached me directly to ask if we could develop an online research methodology course for COSECSA. I wrongly thought it would be easy to change our existing course into an online course and agreed.

Aim of the Course

1. To develop a basic research methodology course for year 1-2 surgical residents in COSECSA countries who have had limited previous exposure to research concepts but who have had basic teaching in epidemiology at undergraduate level.

2. To build on the contents and outcomes of the 2015-16 pilot courses to keep the strong points of the original course but to modify the contents so that a trainee could work through the presentations on their own without the presence of a tutor.

3. To guide course participants to develop their own research ideas into viable projects that will deliver robust clinical evidence.

4. To deliver the course through a viable user-friendly and affordable online platform for interactive learning and chat groups with other participants and online tutors.

5. To deliver a secure online final course test that produces a certificate of successful completion that trainees can include in their training portfolio.

6. To demonstrate educational success to the COSECSA Education and Research Committee and COSECSA Council.

Methods

(a) Course Principles

We reviewed a number of research skills courses available to trainees in different centres, e.g. University of Zambia, South Africa College of Medicine, West Africa College of Surgeons from the USA, United Kingdom surgical colleges. In my opinion some of these courses were too basic and some were too complex or too heavily focused on statistics. We therefore decided to base the course on the following principles:

1. It must explain basic clinical research principles.
2. Use young surgeons/trainees with strong research portfolios as tutors.
3. Use the stepwise approach developed during pilot courses (Introduction, Planning, Reading literature, Methods, Results, Analysis, Presentation, Sustainability).
4. Participants must understand how to use internet search engines and how to read scientific papers critically.

5. Statistics: include enough on statistics that participants understand how statistical analysis and methods work, that they can choose the correct analysis method for their particular data and use online statistics programmes, but not expect them to do complicated statistics themselves.
6. Participants should understand the process of developing a research idea into a functioning study, and do that in practice.
7. Give opportunity for practical sessions/workshops in parallel within the course so that participants can develop their own ideas in to research questions and viable studies.
8. Learn to critically discuss their own and colleagues' research ideas and proposed projects in a constructive and non-confrontational manner.
9. To develop enthusiasm to take research further and to create local clinical research support groups amongst surgical trainees, with mentoring from senior clinicians if possible.

(b) Teaching Methods

The course is set up to progress through a series of progressive steps, as set out in the detailed curriculum. This will take participants from understanding research, through critical reading of the literature, to understanding how to design a study, collect and process data, and finally to present their collected evidence in a meaningful way, with suggestions to maintain enthusiasm. The focus is on clinical research that improves outcomes of care for patients and communities.

In each section there is a PowerPoint lecture or two, followed by suggestions for practical work. The lectures have been modified from the usual to include explanatory notes that a lecturer would usually deliver; all the lectures are therefore such that a course participant can work through these independently. Practical sessions work best if they are done by a few participants together or through online group meetings such as in chat rooms.

For critical evaluation of the literature a number of open access surgical papers are provided in Appendix 1; this gives all participants equal opportunity whether they have easy internet access or not.

In each section there will be a Quiz to help focus learning points. Questions in the Final Course Test will be very similar to the quiz questions.

In Summary

We trust that this online research methodology course (RMC) will help surgical trainees from early in their careers to understand clinical research principles, enable them to read surgical literature widely with a critical mind, enthuse them to develop and complete their own projects with success, and empower them to deliver the evidence that will change local, regional and international clinical practice.

I wish to express my sincere gratitude to Prof Laston Chikoya, Prof Abebe Bekele, Prof Amezene Tadesse and Mr Bob Lane (president IFSC) for their enthusiasm and support from the start of this project all the way through. Our gratitude also goes to the Royal College of Surgeons in Ireland (RCSI) and Dumfries-Devorgilla Rotary Club who made it possible to run the original residential RMC pilot courses. I wish to thank all contributors for allowing me to go back to them again and again for new or altered material. Lastly I wish to especially thank my wife, our children and my oncologist who walked with me through an arduous treatment pathway over the last three years.

J S (Fanus) Dreyer (Editor)
MBChB, MMed(Surg), FCS(SA), FRCSEd, FCS(ECSA), M.Ed(Lond)
Consultant General surgeon, NHS Dumfries & Galloway, Scotland, UK.
Director of Education, International Federation of Surgical Colleges, London, UK.
Honorary Professor in Surgery, Addis Ababa University, Ethiopia.

Dumfries, UK, April 2020

Contributors

David R Ball BA(hons) Oxon, MBBS, FRCA; Retired Consultant Anaesthetist, UK.

Abebe Bekele MD, FCS(ECSA); Professor of Surgery and Dean of Health Sciences, University of Global Health Equity, Rwanda; previously Dean of Medicine, Addis Ababa University, Ethiopia.

Alison Bradley BSc (Hons), MBCHB, MRCSEd, PGC Research Methods; Specialty Registrar in Surgery, West of Scotland, UK

Ainhoa Costas-Chavarri MD, MPH, FACS; Instructor in Surgery, Harvard Medical School; Rwanda Military Hospital, Kigali, Rwanda.

Barend A Dreyer BMSc(hons), MBChB, MRCS, Cert Med Ed (AMEE); Specialty Registrar in Urology, Edinburgh, UK.

Frieda E Dreyer MBChB, MRCOG, Cert Med Ed (AMEE), Specialty Registrar in Obstetrics & Gynaecology, Newcastle-on-Tyne, UK.

Stephan B Dreyer BMSc(hons), MBChB, MRCS, PhD; Clinical Lecturer and Specialty Registrar in Surgery, University of Glasgow, UK.

Vanessa Fawcett MD, MPH, FRCSC; Assistant Professor in Surgery, University of Alberta, Royal Alexandra Hospital, Edmonton, Alberta, Canada.

Jonathan A F Hannay BSc(Hons), MBChB, FRCS, PhD; Consultant Surgeon, NHS Greater Glasgow & Clyde, UK.

Sharukh Sami MBBS, DNB (Gen Surg), MRCS; Specialty Doctor in Surgery, NHS Dumfries & Galloway, UK.

Noel Thin MBBS, FRCS, MD(Res); Specialty Registrar in Surgery, North East Thames, UK.

Suzanne E. Thomson BSc(MedSci), MBChB, MRCS, PhD; Specialty Registrar in Plastic & Reconstructive Surgery, Scotland, UK.

A.3: COSECSA Research Methodology Course (RMC): Detailed course curriculum (2020)

Introduction to the Course

Please read this introduction. It explains the background to the course, its aims, teaching methods and stepwise learning goals, results from previous pilot courses and how this course can help you to take your own research ideas and develop these into a well conducted study and scientific presentation that will add to clinical evidence in surgery.

B: Background: Understanding Clinical Research

B.1: Lecture: Introduction to research principles (JSD):

This lecture will address

1. The difference between research and audit.
2. Different types of research, with focus on clinical research
3. The process of designing a study.
4. Examples of RCTs in Surgery.
5. Data Analysis.
6. Qualitative research, with an example.
7. Surgical Audit
8. Clinical value of research (Why should a surgeon know about this?)
9. Ethics of clinical research

B2: Quiz on principles of research

Quizzes help to imbed essential knowledge and are for personal information only; scores are not collated and do not contribute to you passing or failing the course.

B.3: Practical 1 (share ideas either directly in a discussion group that meet regularly or with your peers in an online group):

Write your own answers to the following questions and then discuss in your newly created study group:

- What do I want to investigate?
- Is it audit or research?
- Is it clinical, laboratory or community based?
- Is it ethical?
- Can I write a research idea in 1-2 sentences?

C: Planning your research

C.1: Lecture: Study design (ST)

This will cover how to select a research topic and title, and how to ask a research question.

C.2: Quiz on study design and formulating the research question

C.2: Practical 2 (individual and in group):

1. Can I write my research idea into a single sentence research question? It is important to move from loose ideas to a structured question in writing that makes sense to anyone else who would read it.
2. Share research ideas and questions in the chat room. Are there others with the same ideas as mine? Can we plan a project together? What are the immediate obstacles and how do we think we can overcome these?
3. Is our local research group strong enough to discuss individual or shared research plans regularly?

D: Collecting Background Information

D.1: Lecture: How to do a literature review (BD, ABe)

Short lecture on internet search engines and demonstration of PubMed and Google Scholar, with additional reference to Hinari (WHO site).

D2. Practical 3 (using PubMed, Google Scholar and Hinari)

Online practical: Go to different search engines and use your own search terms to search for articles in PubMed, Google Scholar and Hinari on topics that interest you. If you already have a good research idea or a potential research question, use this exercise to collect a list of papers you might want to read. (The next lecture will explain to you how to analyse these papers critically and to decide if they are really adding value to your background information and your own research project).

D.3: Lecture: How to read research papers critically: understanding the principles of analysing a paper (JH)

D.4: Practical 4 (critical paper analysis)

A number of scientific papers are made available for reading; these look good at first glance, but might have serious deficiencies. Work through a few papers, review and summarise a few; the summary must address (1) Strong points, (2) Deficiencies, (3)

Will this paper change your clinical practice, and how? It works best if you discuss your ideas either in your local group or in an online chat room (one person summarises one paper but everyone reads all papers under discussion); then try to form a group opinion and test it against other groups. You will find there are no fixed rules; different readers interpret and value the same paper differently but you can come to shared conclusions by consensus. The next step is to find papers in the literature according to your own speciality or current knowledge, and analyse these critically, but always write down your opinion on points (1), (2) and (3) as above.

Please see Appendix 1:

Appendix 1 contains a number of journal papers for critical reading and analysis. These papers are all available for free download through e.g. Google Scholar and copies are available in the course handbook. It works best if 1-2 persons read and summarise one paper and then present their findings to colleagues within a study group. If you give 10 minutes for each presentation and 5 minutes for discussion you can go through 5 papers in under 90 minutes. That should give you sufficient confidence to be a critical future reader.

D.5: Quiz on planning and preparation.

E: Methods

E.1: Collecting and managing your project data (lecture) (ABr)

E.2: Quiz on methods and study design

E.3: Practical (individual and in small groups)

Aim is that you and/or your group will now be able to write down a study proposal with appropriate research title, research question, study design, planned methods. Also discuss specific practical problems in research methods that you have encountered or envisage.

Start to ask yourself:

- What is your research question? Is it robust and unambiguous?
- How to select the correct study design for your research question
- What research method(s) will work best for your question?
- End points, variables, feasibility?
- How will data be collected and who will do it?
- What is the desired and/or available time frame to complete data collection?
- How will data be analysed and by whom?
- How to prepare for ethical approval for your study.
- Time management.

- Recruiting other team members.
- Reviewing end-points regularly.

F: Results: Analysing your data

F.1: Lecture: Statistics for Medicine (Understanding basic principles of statistics) (NT,VF)

This lecture starts by explaining the basic principles of statistics to help understand the principles of chance, how that is used to calculate statistical chances and use these to interpret the spread of data, and to predict transferability of study results to wider populations. It explains different techniques and tools of statistical analysis , summarise more advanced statistical methods and explain how statistical calculations can make practical/clinical interpretation and implications of research results more understandable and transferable, but also how excessive statistical tests can overcomplicate result interpretation.

F.2: Practical 6 (using online statistics programmes)

Find and select a statistics programme that is available for free online use (open access) through an internet search engine (e.g. Google). There are many programmes (e.g. JASP, SOFA, GNU PSPP, Jamovi, IBM SPSS, MacAnova, Invivostat, etc etc) but they all have different strong and weaker points. Sites like GoodFirms (goodfirms.co) try to summarise these (e.g. IBM SPSS is very powerful, MacAnova also works with Linux, Invivostat identifies and removes inaccurate data). Try different programmes with your own experimental data and see which gives you the best understanding of the data and the best visual representation to share your findings (do not try to find a programme to compensate for a poorly conducted study).

Take either data from your own provisional results, or from a pilot study, or take the data from a robust paper you have read, and enter the data. Best is if 2-3 people use the same data in more than one online programme and compare outcomes, then make sure you understand the way the programme does the statistics (you only need to understand the principles of how the quantitative data becomes readable statistics), then discuss which programme would be best suited to your own project.

F.3: Quiz on statistics

F.4: Lecture: Qualitative data analysis (JSD)

F.5: Extra dimensions of data analysis (JSD)

a. **Quality of Life indicators**
b. **Patient satisfaction outcomes**
c. **Quality control/Patient Safety factors**

F.6: Quiz on data analysis

F.7: Practical (how to analyse data from your own study)

Discuss potential analysis methods for study proposals in your group e.g. what type of data (quantitative, qualitative) do we expect to get?, therefore what tests to use?; can we find a statistician to help?; which internet available statistics programme can we use? do we need to quantify any qualitative data? Let the principal investigator for each study present their ideas for 5-10 minutes, and then have a 10-15 minute discussion; end by writing down 3 key points on data analysis for each study proposal.

G: Reporting your Findings

G.1: Lecture: Presenting my data (FED)

How to plan and complete Podium (oral) and/or Poster presentations of your project results in your local department/hospital, at scientific meetings/conferences or to specific interest groups (e.g. Dept of Health, Health care managers).

G.2: Lecture: Principles of scientific writing (SD)

How to plan, construct and write the detail of a scientific paper, how to choose a journal for publication and how to give yourself the best chance of getting your paper accepted for publication. How to deal with paper rejection and try again.

G.3: Quiz on presentations and writing principles

G.4: Practical 8 (presenting and writing)

You can practice presentation and writing before you have completed your own research project as long as you have a small group of research enthusiasts that support each other. Things you can do include:

a. Prepare an oral presentation on a paper you have read as if it were your own study and you have to share the results with your peers.
b. Prepare an electronic poster on a paper and present at your meeting (you can do the poster in PowerPoint and project from your laptop, so no expenses necessary).
c. When you have read 5-10 good articles as background for your study proposal, summarise these into an oral presentation to your study group or your own hospital department. You can summarise these under the IMRD headings you will use to write your own paper.
d. Write a summary abstract for yourself on these background papers, as if you are going to submit this as an abstract for a conference.

e. Use (c) and (d) as the core information to write up the literature review for when you want to publish your own study.
f. Write up a summary of the methods you want to use for your own study and present to your peer study group. Welcome any positive criticism to improve your project.

H: How do I keep going?

H.1: Lecture: Research with limited resources; research collaborations; how can I make research work in my own institution? (ACC)

H.2: Practical 9 (the value of a local research support group)

Try to find likeminded colleagues who are your peers, e.g. other surgical residents that you work with in the same hospital, or with residents from other departments, or other health care practitioners that are interested in clinical research (laboratory staff, nutritionists, public health doctors [who often know more about statistics and epidemiology than surgeons] etc). The trick is to think widely. Invite 1-3 supportive consultants to attend, even if they might be very senior academics. Most professors love young researchers to come up with new ideas and would like to know about these early. Find a time and place to meet regularly when it would suit everybody to meet most of the time. Do not overdo the meetings; monthly meetings will be more sustainable than starting with weekly meetings because you are enthusiastic. Keep very short minutes of you meetings so you can remind people what you discussed previously and have an agenda for every meeting, so people can prepare to discuss their specific project successes or problems. Invite speakers to come and talk about difficult issues e.g. a statistician from the local university (irrespective whether medical or not), somebody who has achieved research success in a different field, someone who can advise on publication (e.g. an editorial board member of a medical journal). Most important is to be open, supportive, non-threatening and non-defensive. Criticism must always be positive and aimed at the contents the group discusses; never be critical of researchers in person, even if they do not attend the meeting. Make every meeting a learning experience and end the meeting with a short list of 3-5 learning points from that specific meeting. Then share a meal or a few drinks or go home if you have a family and don't talk about research when you have left the meeting.

J: Final Course TEST

This is a final Pass or Fail course consisting of 25 questions worth 4 points each. Each question has a choice of 5 potential answers: if you get it right at first attempt you get 4 points, if wrong and you try again, you will get 3, then 2, 1 and 0. You have to score 80% to get a pass mark. You can take the test ONCE only, so make sure

you have mastered the course content and done the practical exercises. If you do not pass you will have to reregister to repeat the course.

The TEST will be developed after we have come through the Covid-19 pandemic.

K: Collecting your Certificate

This will be an electronic certificate issued through the COSECSA School for Surgeons. You can save the pdf copy to your electronic portfolio or print a paper copy.

B.1: Lecture

Understanding the Principles of Surgical Research and Audit
(including Ethics of clinical research)

J S (Fanus) Dreyer

1

This lecture will address:

1. The difference between research and audit.
2. Different types of clinical research.
3. The process of designing a study.
4. Examples of RCTs in Surgery.
5. Data Analysis.
6. Qualitative research, with an example.
7. Surgical Audit
8. Something about Ethics in research

2

What is Research and what is Audit?

- Audit = ask “what happens”?
- Research = ask “why” or “how” does it happen

3

SURGICAL RESEARCH

4

Research

A: Different types of Research:

- Case study (=one case that brings new learning)
- Clinical series (=a number of similar cases that describe something new)
- Observational Study (=observing a large group over time [=longitudinal] e.g. the Framingham study)
- Cohort studies (=comparing two groups of people, usually quite similar except for the research intervention)
- Uncontrolled trials
  - Retrospective trials (=looking back at treatment effect afterwards)
  - Historical controls (=comparing current results with how things used to be)
  - Non-Randomised studies (=two study groups that might look similar but patients in each group selected with inherent bias)

5

- Controlled Trials:
  - Prospective (= recruit patients only after study design has been completed)
  - Randomised Controlled Trial (RCT) [=patients allocated without choice e.g. patients draw numbers and all equal numbers go into treatment arm A and all unequal numbers go into treatment arm B]
  - Blinded studies (Double blinded) [=the clinical researcher (patient doctor) and the data analyser (e.g. statistician) do not know which group received which treatment until after the reults are reported]
- Meta-analysis (=combining the results of different similar studies to get a stronger combined result – see later)
- Qualitative Research (see later)
- Surgical/Clinical Audit (see later)

6

Definition of Clinical Research:

Clinical research intends to produce knowledge valuable for understanding human disease, preventing and treating illness, and promoting health. It involves interaction with patients, diagnostic clinical data or populations.

Penson DF & Wei JT
Clinical Research Methods for Surgeons
2006 Humana Press

7

How do we use Evidence from research?

- Level 1: High quality meta-analyses (1a);
  Review of RCTs;
  Well designed RCTs (1c).
- Level 2: High quality cohort studies or case control studies;
  or Review of such studies.
- Level 3: Non-analytic studies, e.g. Case studies or series.
- Level 4: Expert opinion

8

To do a RCT in Surgery is very difficult (often almost impossible):

An example of a RCT in Colorectal Surgery that greatly influenced my practice:

Dutch TME trial

(Look up the technical aspects and oncological benefit of Total Mesorectal Excision (TME) for rectal cancer if you are unfamiliar with the procedure)

9

Total Mesorectal Excision (TME)

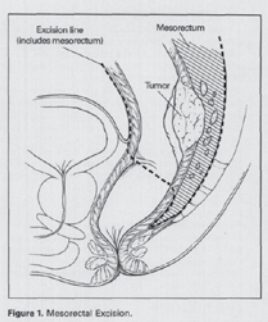
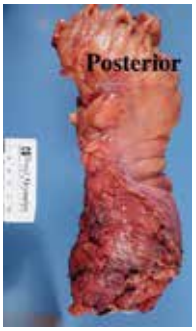


Figure 1. Mesorectal Excision.



10

Dutch TME trial

The NEW ENGLAND JOURNAL of MEDICINE

Ellen Kapiteijn, Corrie A M Marijnen, Iris D Nagtegaal, Hein Putter, et al.

Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer

NEJM Aug 30, 2001; 345 (9): p. 638-646

11

Dutch TME trial

This study looked at the outcomes for general colorectal surgeons who learn to do TME correctly, with and without pre-operative radiotherapy; the surgeons were not rectal cancer experts in academic centres but general hospital colorectal surgeons who do rectal cancer surgery as part of wide colorectal practice. Because the benefit of radiotherapy in this cohort was not yet established it was ethically acceptable to do a trial of with or without pre-op radiotherapy.

12

The study showed that for Stage 3 cancer the local recurrence rate with pre-op radiotherapy was a third of those who did not have radiotherapy (and for stage 2 cancer it was 1% vs 5.7%), so in daily rectal cancer surgery the radiotherapy brought significant benefit.

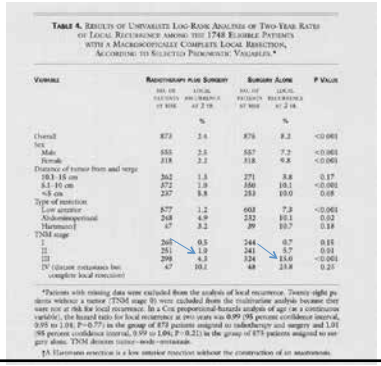


TABLE 4. RESULTS OF UNIVARIATE LOG-RANK ANALYSIS OF TWO-YEAR RATES OF LOCAL RECURRENCE AMONG THE 1748 ELIGIBLE PATIENTS WITH A MACROSCOPICALLY COMPLETE LOCAL RESECTION, ACCORDING TO SELECTED PROGNOSTIC VARIABLES.*

VARIABLE	RADIOTHERAPY PLUS SURGERY		SURGERY ALONE		P VALUE
	NO. OF PATIENTS AT RISK	LOCAL RECURRENCE AT 2 YR	NO. OF PATIENTS AT RISK	LOCAL RECURRENCE AT 2 YR	
		%		%	
Overall	873	2.4	875	8.2	<0.001
Sex					
Male	555	2.5	557	7.2	<0.001
Female	318	2.2	318	9.8	<0.001
Distance of tumor from anal verge					
10.1–15 cm	262	1.3	271	3.8	0.17
5.1–10 cm	372	1.9	350	10.1	<0.001
≤5 cm	237	3.8	253	10.0	0.01
Type of resection					
Low anterior	577	1.2	602	7.3	<0.001
Abdominoperineal	248	4.9	232	10.1	0.02
Hartmann†	47	3.2	39	10.7	0.18
TNM stage					
I	265	0.5	244	0.7	0.95
II	251	1.0	241	5.7	0.01
III	299	4.3	324	13.0	<0.001
IV (distant metastasis but complete local resection)	47	10.1	48	23.8	0.25

*Patients with missing data were excluded from the analysis of local recurrence. Twenty-eight patients without a tumor (TNM stage 0) were excluded from the multivariate analysis because they were not at risk for local recurrence. In a Cox proportional-hazards analysis of age (as a continuous variable), the hazard ratio for local recurrence at two years was 0.99 (98 percent confidence interval, 0.95 to 1.04; P=0.77) in the group of 873 patients assigned to radiotherapy and surgery and 1.01 (98 percent confidence interval, 0.99 to 1.06; P = 0.21) in the group of 875 patients assigned to surgery alone. TNM denotes tumor–node–metastasis.

†A Hartmann resection is a low anterior resection without the construction of an anastomosis.

13

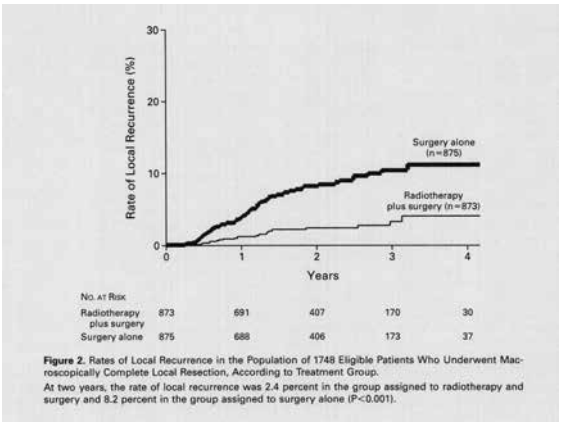


Figure 2. Rates of Local Recurrence in the Population of 1748 Eligible Patients Who Underwent Macroscopically Complete Local Resection, According to Treatment Group.

At two years, the rate of local recurrence was 2.4 percent in the group assigned to radiotherapy and surgery and 8.2 percent in the group assigned to surgery alone (P<0.001).

14

Meta-analysis

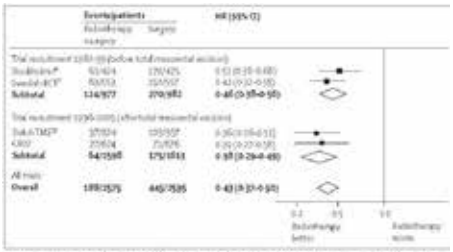


Figure 3: Summary of reduction in risk of local recurrence in phase III trials that have assessed short-course preoperative radiotherapy with 5 Gy per fraction

In this meta-analysis the data of the Dutch TME trial was combined with the UK CR07 trial that also looked at pre-op radiotherapy with TME and with older studies from Sweden that looked at pre-op radiotherapy with classic rectal cancer surgery (before TME). All studies showed benefit for pre-op radiotherapy.

15

How do we plan a study?

PRINCIPLES:

1. Inductive method (from Newton):
   "Let the data speak for itself"

(cf. Deductive method: used in mathematics, astronomy & physics, where a theory is tested against observations)

16

2. Structure of the research plan:

"IMRD" format (from Pasteur)

= Introduction,
Methods,
Results,
Discussion.

17

How to Write a Research Proposal:

1. Title: -Must have focus ("What are you working on?")
   -Will evolve
   -Not the same as manuscript title

2. Contributors: -Surgeons
   -Scientists
   -Data management experts

18

3. Research Question:

- This is the most important part of the research plan.
- Ask: "What do I want to find out?"

(a) Types of Research Questions:
   -Is it true?
   -What is the truth?
   -Is it better?
   -What do we not know?

(b) What is the significance?
   -Must have clinical relevance, i.e. Make patient care better.

19

4. Background knowledge:

- Study both successful and unsuccessful research in your field of interest.
- Read relevant textbooks and journal papers.
- Discuss with colleagues.
- End with small group (about 10) of key references.

20

5. Define a Study Group:

Consider future patient characteristics when designing your study.

-Inclusion criteria
-Exclusion criteria
-Comparison groups
-Time frame for recruitment

21

6. End Points:

(a) What outcome(s) do you want to measure?

E.g: -Mortality
-Morbidity:
e.g. -Short term: Post-operative Complications
-Medium term: Cancer recurrence
-Long term: Disability
-Health related QoL
-Patient Satisfaction
-Quality of Care (e.g. Patient safety)

(b) Each end point must contribute to answering the research question.

22

7. Variables:

- = different factors that might influence results.
- Can be confusing and distracting
- Solve this by organising these medically, e.g:
  - Demographics
  - Classify clinical conditions (e.g. ASA grade)
  - Morphology (variations of anatomy or pathology)
  - Co-morbidity
  - Surgical details
  - Support mechanisms

23

8. Data analysis: Who? How?

9. Feasibility: (Potential Obstacles)

- Sample size & Comparison group
- Timetable
- Limitations & Problems
- Ethics
- Budget

10. Review of the Proposal:

- Academic head/Educational supervisor
- Ethics committee
- Institution's review board/Ethics committee

24

BASIC STATISTICS

- Statistics = a method to extrapolate data from samples to populations.
- Statistics is NOT an exact science.

(a) Descriptors of the Centre:

-Mean
-Median
-Mode

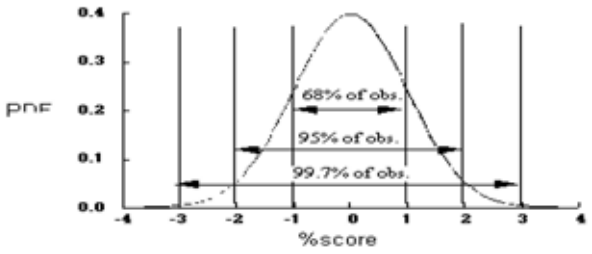
(b) Descriptors of the Spread:

-Variance
-Standard Deviation
-Percentiles

25

Distribution and Symmetry:

Normal distribution =



26

Quantifying Uncertainty: THE *p* VALUE

- What is a statistically significant *p* value?
- What does a *p* value of $\leq$ 0.05 mean?
- A statistically significant *p* value does not mean there is a clinical significance between groups.
  *Why? (Answer on next slide)
- The effect of sample size on *p* value.
  **Type II error.

27

Some explanation

*If you test two drugs in thousands of people one might be marginally better than the other; this will be statistically significant but does not make any clinical difference if patients don't like the taste of the better drug.

**Try to work out how very small sample sizes can give you false results. Type 1 and 2 errors are explained in detail in Statistics lectures.

28

Statistical Tests to compare samples:

- Parametric test (for data with a normal distribution)
  = *t-test*
- Non-parametric data (for asymmetrical distribution)
  = Wilcoxon-Mann-Whitney test
- To compare outcomes: *Chi-squared* test
- For multiple complex variables: Linear and Logistic Regression analysis

29

Qualitative Research

Definition: What is Qualitative Research?

= methods to understand human experience of illness, health and treatments.

=How do patients or health care workers find meaning in illness and health, and how it affects their work, life, relationships etc.

30

Qualitative Methods:

1. Interviews
2. Group opinions
3. Questionnaires
4. Case Studies

Data analysis:

1. Test the Validity and Reliability of the data, e.g. through Triangulation.
2. Cronbach alpha is used to try and quantify qualitative results.

31

SURGICAL AUDIT

What is Audit?

32

What is Clinical Audit?

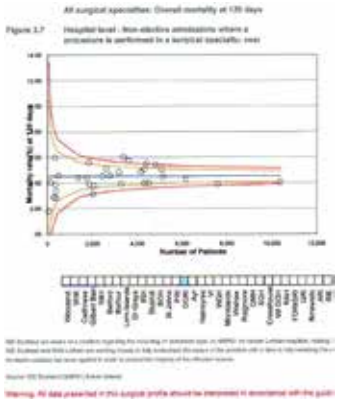
"A process that improves the quality of patient care through systematic review of care against explicit criteria and supports changes in practice to meet those criteria."

What is Audit as a Research technique?

"Assessing practice honestly enough to notice differences in outcome and report such data."

33

Examples of Surgical Audit: QiS Surgical Profiles

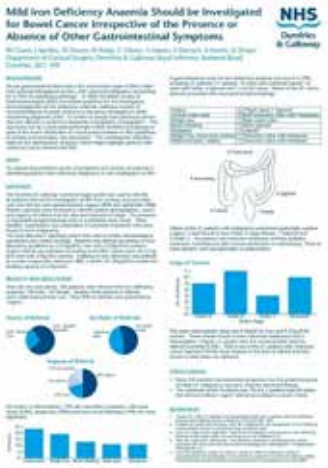


34

Explanation of previous slide

The graph looks at 120 day mortality after all surgery in Scottish hospitals. Low volume hospitals are on the left with more variation of results allowed (because one extra death in a small number of patients will increase your mortality rate a lot more). The yellow and red lines are for 1 and 2 standard deviation lines, so 68% of results are within the yellow line and 95% inside the red line; above the red line means that a hospital's mortality is worse than 95% of others, so that then needs to be investigated.

35



This was a simple audit of asking whether patients with mild iron deficiency anaemia can have bowel cancer, and the answer was that sufficient numbers of women in a rural Scotland setting can have bowel cancer when they present with mild anaemia and no other symptoms. The study simply asked What is happening and did not try to explain Why.

36


Personal Bedside Patient Televisions Might Increase the Risk for Deep Vein Thrombosis

NHS Dumfries & Galloway

Again a simple study that looked at how many in-patients in surgical wards in one hospital were diagnosed with DVTs a year before and after bedside TVs were introduced. It does not try to explain why and, because the numbers are low, it does not try to present this as statistically significant. It is presented just to make hospital users aware of a higher DVT incidence. In depth research can then find out the Why and How.

37

I encourage surgical trainees to do Clinical Audit:

1. It is cheap.
2. It does not need ethics approval (you still need to write a proposal).
3. It is based on clinical work.
4. It improves practice and patient care.
5. It makes you used to analysing clinical outcomes and reporting these honestly.
6. It takes you away from a "blame culture" for mistakes and teaches you to analyse the cause of errors.
7. It can improve patient safety.

38

Audit in Safe Surgery:

(a) Audit the process of using a Safe Surgery Checklist (SSC), like the WHO checklist:

e.g: -how the SSC is implemented and/or accepted
-practical problems in using different parts of the SSC
-developing a protocol for swab and sharps counts.

(b) The effects of the SSC on Specific Outcomes (through quantitative/statistical research)

e.g: -wound infection or DVT incidence
-drug allergic reaction incidence
-delay in getting blood to theatre

(c) Qualitative research on how the SSC affects practice:

e.g: -how did introducing the SSC affect theatre teamwork or communication.
-how difficult was it to change theatre practice.

39

Clinical Audit is thus a Simple Link between Surgical Research and Safe Surgery.

40

Ethics in Surgical Research

History of non-ethical research:

1. 1946 Nuremberg doctors trial
2. 1960s Thalidomide tragedy
3. 1932-1973 Tuskegee syphilis study

"The slightest hint of financial or material conflict of interest will jeopardize your credibility as a researcher"

41

1. Nuremberg Code
2. Declaration of Helsinki 1964

- Human subjects in research must give informed consent
- The expected benefits should justify the research and risks
- Human studies should be based on animal studies and knowledge of the natural history of the condition
- Physical and mental suffering and injury must be avoided
- Subjects should be able to withdraw from the study at any point
- Studies should be conducted by qualified scientific personnel who will be prepared to terminate the study at any point if they think subjects could be at significant risk

42

Informed Consent in Research

Must have three qualities:

1. Information
   - What are you going to do?
   - What are you not going to do?
2. Understanding
   - Must have somebody outside the study team to talk to
3. Voluntary agreement
   - Patients who are very ill or injured or going for major surgery are traumatised. It is not a good time to make decisions. They must NOT be put under more stress.

43

Required elements of informed consent in research [1]

1. State that the intervention is research.
2. Purpose of the research
3. Describe study procedures
4. Potential risks
5. Potential benefits of participation
6. Alternative treatments
7. Methods to maintain confidentiality
8. Number of subjects in the study

44

Required elements of informed consent in research [2]

9. Compensation for harm
10. Investigator contact information
11. Statement that participation is voluntary
12. Statement that there may be unforeseen risks
13. Reasons why a participant may have to be removed from the study
14. Any extra costs for participation
15. Adverse effects for early withdrawal
16. Explain how results will be reported

45

Summary of Ethics in Clinical Research

1. Non-maleficence
   - First do no harm.
2. Beneficence
   - Excellence
   - Accountability
   - Humanism
   - Altruism

46

Questions?

Write your questions down so that you can discuss with colleagues or ask your tutor.

47

B2: Practical 1: Finding a research idea

Ask yourself the following questions

- What do I want to investigate?
- Is it audit or research?
- Is it clinical, laboratory or community based?
- Is it ethical?
- Can I write a research idea in 1-2 sentences?

Now meet with your colleagues and then ask help from a mentor

Each participant presents a research idea that they should have written down in advance (2 min) and the group discusses, e.g. is it research or audit? what would the research question be? how easy or difficult will it be to do? The mentor gives guidance and then you decide either as an individual or as a group which idea you are going to develop further into a research project in parallel to completing the research methodology online course.

C.1: Lecture

Planning your research: Study Design

Suzanne E Thomson

1

This presentation will address:

- A brief overview of study designs
- Selecting a research topic and title
- Asking the right question
- Worked examples
- Further resources

2

Notes:

High quality research requires a reliable and robust study design to answer a well formulated question in the most efficient and accurate way possible. Good information available at

http://research.library.gsu.edu/c.php?g=115595&p=755213

3

Hierarchy of Research Design

Different designs for different questions.

4

Notes

High-quality evidence begins with a suitable research design. The figure shows the basic hierarchy of clinical study designs. The "weakest" form of evidence comes from single case reports. These are the anecdotal reports of the outcomes seen in 1 or 2 patients but may still be very useful in describing a rare condition or pitfall. The strongest type of evidence comes from meta-analyses and randomized controlled trials (RCT) that enrolled enough subjects so that the results have meaning.
Keep in mind that we don't always need an RCT to determine with reasonable confidence whether a health technology works and is safe. Sometimes other types of studies provide high-quality evidence as long as they are well designed, well executed, and applicable to the patient population in which we're interested. Moreover, even the best study can be fatally flawed if it's poorly executed. Designing your research to succeed takes into account many factors . . .

5

Research Design Definitions

- **Meta-analysis** – statistical analysis that combines data from several studies
- **Systematic review** – critical evaluation of available literature, may include meta-analysis
- **Randomised control trial** – assigns participants to one or more treatment arms through a defined method of randomisation to reduce bias

6

Research Design Definitions

- **Cross-over study**
- **Cohort study** (prospective observational study) – participants are in predefined groups and are observed for specific outcome measurements over a defined study period
- **Cross-sectional study and surveys**
- **Case series or report**
- **Ideas, Editorials, Opinions**

7

Notes

Cross over study is when a treatment is give for a set time, followed by a washout period in the case of drug trials and then a different treatment to the same group – in this way individuals act as their own control.
Cohort studies observe groups of individuals before they develop a disease or a particular outcome.
Cohort studies have the power to detect many different outcomes of an exposure and allow researchers to calculate a relative risk of developing a disease based on different exposures.
It may take many years to detect changes in the groups.
Because of the time involved and number of participants needed, cohort studies may be very costly.
Surveys gather data to describe the **demographics** of a group; the **health status** of a group of people at a particular time; the utilization of medical services; or the knowledge, beliefs, and attitudes of people regarding health practices.
Surveys are a major data collection method in health services research.
Survey research is extremely complex.
Survey results often are difficult to interpret and generalize to other groups and time periods but provide wonderful insights into the practices and health conditions of large groups of people as well as clues for future investigation.

8

Your Research Design Depends on Your Question

- Work backwards
- What do you want to study - - -> how will you most accurately assess this
- Select representative outcome measures
- Adequately power your study
- Appropriate statistics
- Eliminate bias
- Pragmatism

9

Selecting a research topic

- Something **you** are ***interested*** in
- Something important to your patient group
- Consider local expertise
- Review the available literature
- Generate the correct Research Question

10

Asking the correct question

- Vital to the rest of your research process
- Irrespective of type of research
- Adequately define and refine

- The right question helps focus your efforts during the research process

11

Notes

Getting the question right is absolutely crucial to the rest of the research process. This is obvious as getting the right answer to the wrong question is going to waste an awful lot of time and resources. However, the main problem is not asking the wrong question, but not properly defining the right question. The key to defining an RQ is focus. The end product needs to be a specific query that is explicit in what it is looking for. The process of defining the question is therefore essentially one of taking a broad topic area and narrowing it down until you have a question that can be answered fully.

12

Refining your research question

- What do you want to find out?
- Can you answer the question within the time and resources available to you?
- What outcome measures will you use?
- What participants will you use? How will you access them?
- Are there any ethical issues?
- What will make your research stand out/interesting/ useful?
- What impact will it have and how will it effect change?

13

Notes

Often you can find many interesting topics during a days work. Given all the resources and time in the world you could plan a detailed study to compare the sensitivities of every test available to investigate non specific abdominal pain in all patient groups, however, you and I have neither the resources or time even if we have the inclination! You need to add some pragmatism and decide where to focus your energies and this will take into account several factors. Firstly it is important to you and your patients. Then once your practice throws up a topic decide what exactly is the problem and the potential solutions – e.g. too many skin grafts failing, and you think it has to do with the dressing. Then you think that you would like to used different dressings. Then you need to define which ones. How will you know if your intervention has worked? Which participants will you use? How will you access them? What controls will you use and how will your findings impact and effect change locally/globally?

14

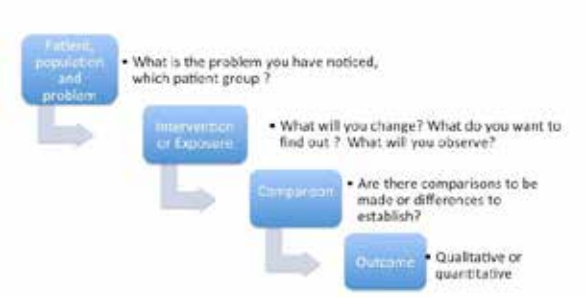
ReFINER

Is it…

- **F**: Feasible -- adequate caseload, time, money.
- **I**: Interesting
- **N**: Novel -- thorough understanding of available literature – and where your work fits in
- **E**: Ethical
- **R**: Relevant -- to your patients, to general scientific community

15

PICO – a tool for refinement



16

Notes

Once you have refined your area of focus, you need to put the meat on the bones of your research topic and generate a precise question. Begin by defining your population – is it all patients presenting with a particular condition or only those at high risk due to another definable feature e.g. age,. This will allow you to start defining your inclusion and exclusion criteria. What is the intervention or exposure of interest – this may be something you observe in a case series or something you purposefully control in a controlled cohort study. Will you look to compare to another group, and what outcomes will you use to compare – will these be qualitative (e.g. in patient reported measures or surveys of opinion) or will they be quantitative and which scale of measurement will be used? This process will occur without thinking.

T may be added on the end for **TIMING**.

17

Hypothesis driven research

- What do you think will happen (alternative hypothesis)
- Null hypothesis
  - 2 sided – "no relationship or no difference"
  - 1 sided – predict x better than y
- Statistical testing
  - Sample
  - Selected statistic
  - Can you reject null hypothesis

18

Notes

The primary research question should be driven by the hypothesis rather than the data.1,2 That is, the research question and hypothesis should be developed before the start of the study. This sounds intuitive; however, if we take, for example, a database of information, it is potentially possible to perform multiple statistical comparisons of groups within the database to find a statistically significant association. This could then lead one to work backward from the data and develop the "question." This is counterintuitive to the process because the question is asked specifically to then find the answer, thus collecting data along the way (i.e., in a prospective manner). Multiple statistical testing of associations from data previously collected could potentially lead to spuriously positive findings of association through chance alone.2 Therefore, a good hypothesis must be based on a good research question at the start of a trial and, indeed, drive data collection for the study…

19

…more Notes

The research or clinical hypothesis is developed from the research question and then the main elements of the study — sampling strategy, intervention (if applicable), comparison and outcome variables — are summarized in a form that establishes the basis for testing, statistical and ultimately clinical significance. For example – I think that using Inadine dressings will improve split thickness skin graft take compared to simple gelonet dressing. However, when formally testing statistical significance, the hypothesis should be stated as a "null" hypothesis. The null hypothesis would be that there is **no difference** in skin graft take between the Inadine and Gelonet dressing groups. Statistics can then be used to assess if there is a significant difference. If the findings of the study are not statistically significant (i.e., there is no difference in functional outcome between the groups in a statistical sense), we cannot reject the null hypothesis, whereas if the findings were significant, we can reject the null hypothesis and accept the alternate hypothesis (i.e., there is a difference in mean functional outcome between the study groups), errors in testing notwithstanding. In other words, hypothesis testing confirms or refutes the statement that the observed findings did not occur by chance alone but rather occurred because there was a true difference in outcomes between these surgical procedures. The concept of statistical hypothesis testing is complex, and the details are beyond the scope of this article.

20

Type I and II Errors

Scenario: Man on trial for stealing Haile Gebrselassie's extensive medal collection

Hypothesis: "The evidence produced before the court proves that this man is guilty."

Null hypothesis (H0): "This man is innocent."

Four potential outcomes . . .

	NULL HYPOTHESIS is valid: Innocent	NULL HYPOTHESIS in invalid: Guilty
Reject H_0 I think he is guilty	Type I error FALSE POSITIVE Innocent man convicted!	Correct outcome TRUE POSITIVE Convicted, justice!
Don't reject H_0 I think he is not guilty	Correct Outcome TRUE NEGATIVE Good man gone free!	Type II error FALSE NEGATIVE Criminal wrongly freed

21

Notes: an Example from elsewhere in life…

Hypothesis: "The evidence produced before the court proves that this man is guilty."
Null hypothesis (H_0): "This man is innocent."
A type I error occurs when convicting an innocent person (a miscarriage of justice).
A type II errors occurs when letting a guilty person go free (an error of impunity)
A positive correct outcome occurs when convicting a guilty person.
A negative correct outcome occurs when letting an innocent person go free.

22

P-value

The probability of rejecting the null hypothesis when it is actually true – i.e. a type I error i.e. detecting a difference when there is none. NT will discuss more.

$P < 0.05$ – means less than 5% chance of rejecting the null hypothesis when it is actually true

	NULL HYPOTHESIS is valid	NULL HYPOTHESIS in invalid
Reject H_0	Type I error FALSE POSITIVE	Correct outcome TRUE POSITIVE
Don't reject H_0	Correct Outcome TRUE NEGATIVE	Type II error FALSE NEGATIVE

23

Refine then Revisit your Question

- Write down your question
- Read more
- Discuss with peers
- Make sure you are still happy with your hypothesis and question
- Don't be afraid to change it

24

Notes

Once you have defined your question you can use this to focus your research. Write it down, read more around the topic, ensure this has not been addressed already in the literature or if it has consider how you will conduct this to make it relevant to your patient population. Discuss the question with peers and revisit the question in a few days to ensure it is the correct one. A solid research question is the basis for non-bias and fruitful research.

25

Clinical equipoise

- A clinical or surgical trial is only ethical if the expert community is uncertain about the relative therapeutic merits of the experimental and control groups being evaluated.

26

Notes

The research hypothesis should be stated at the beginning of the study to guide the objectives for research. Whereas the investigators may state the hypothesis as being 1-sided (there is an improvement with treatment), the study and investigators must adhere to the concept of clinical equipoise. According to this principle, a clinical (or surgical) trial is ethical only if the expert community is uncertain about the relative therapeutic merits of the experimental and control groups being evaluated. It means there must exist an honest and professional disagreement among expert clinicians about the preferred treatment. An ethical dilemma arises in a clinical trial when the investigator(s) begin to believe that the treatment or intervention administered in one arm of the trial is significantly outperforming the other arms. A trial should begin with a null hypothesis, and there should exist no decisive evidence that the intervention or drug being tested will be superior to existing treatments or effective at all. As the trial progresses, the findings may provide sufficient evidence to convince the investigator of the intervention or drug's efficacy. Once a certain threshold of evidence is passed, there is no longer genuine uncertainty about the most beneficial treatment, so there is an ethical imperative for the investigator to provide the superior intervention to all participants. Ethicists contest the location of this evidentiary threshold, with some suggesting that investigators should only continue the study until they are convinced that one of the treatments is better, and with others arguing that the study should continue until the evidence convinces the entire expert medical community.

27

Research title

- Brief
- Accurate
- Reflect content of work
- Enticing

28

Summary

- Decide a topic
- Literature review/expert advice
- FINER
- PICO
- Hypothesis led research
- Clinical equipoise
- Refine question – focused, unbiased research

29

Notes

A poorly devised research question may affect the choice of study design, potentially lead to futile situations and, thus, hamper the chance of determining anything of clinical significance, which will then affect the potential for publication. Without devoting appropriate resources to developing the research question, the quality of the study and subsequent results may be compromised. During the initial stages of any research study, it is therefore imperative to formulate a research question that is both clinically relevant and answerable.

30

Remember the 6 P's

Proper planning and preparation prevents poor performance!

31

Resources

- http://www.nlm.nih.gov/nichsr/ihcm
- Farrugia P *et.* al **Research questions, hypotheses and objectives.** Can J Surg. 2010 Aug; 53(4): 278–281.

32

C2: Practical 2: Asking the research question

1. Can I write my research idea into a single sentence research question? It is important to move from loose ideas to a structured question in writing that makes sense to anyone else who would read it.

2. Share research ideas and questions in the chat room. Are there others with the same ideas as mine? Can we plan a project together? What are the immediate obstacles and how do we think we can overcome these?

3. Can we form a local or research group that discuss a shared or each other's research plans regularly.

Each group write their research idea into a research question. They must decide what is the null hypothesis and the alternative hypothesis for their question. One member of each group functions as a scribe and he/she must record all important conversation points and final decisions. All differences of opinion within groups must be settled by consensus and/or voting within the group.

D.1: Lecture

How to do a Literature search

Barend Dreyer

Abebe Bekele

1

1

Focus on

- Pubmed
  - www.ncbi.nlm.nih.gov/pubmed/
- Google scholar
  - scholar.google.com

2

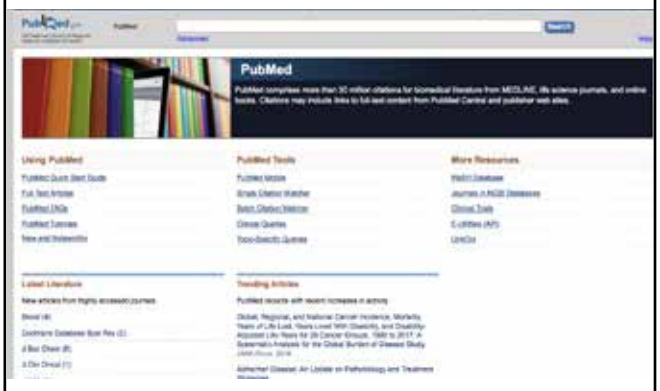
2

Pubmed

- Pubmed is a search engine for life science and medical papers/abstracts from a variety of sources
- Main source is MEDLINE
- Over 5,400 biomedical journals are covered
- Good: comprehensive search technology, access pre-publication abstracts/papers, can track authors and specific topics for developments
- Bad: many papers require paid/institutional access, validity and quality of certain papers may be questioned

3

3



4

4

Example

- I want to know more about "schistosomiasis" as a cause of "bladder cancer"

5

5

Building the Search

1) Decide on topic and think of key words
2) Try to be specific
3) Decide on AIM of search
   - If for systematic review, focus on clinical trials/major studies
   - If for clinical research, focus on review articles

6

6

7

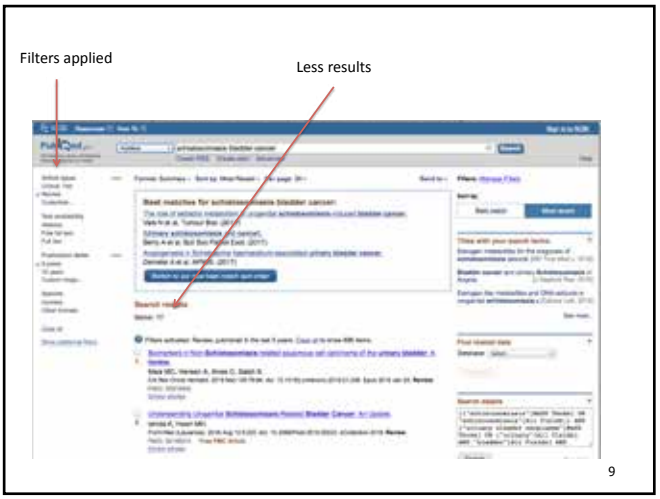
7

Building the Search

1) Decide on topic and think of key words
2) Try to be specific
3) Decide on AIM of search
   - If for systematic review focus on clinical trials/major studies
   - If for clinical research focus on review articles
4) Narrow results using filters
   - Recent studies only
   - Use "review" articles to get outline

8

8



9

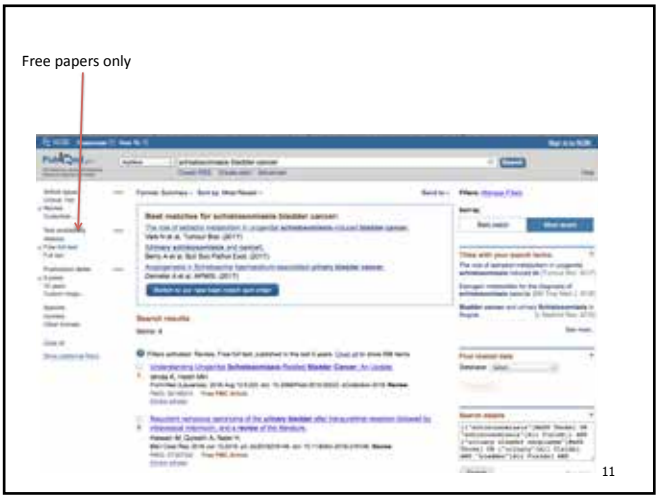
9

Building the Search

1) Decide on topic and think of key words
2) Try to be specific
3) Decide on AIM of search
   - If for systematic review focus on clinical trials/major studies
   - If for clinical research focus on review articles
4) Narrow results using filters
   - Recent studies only
   - Use "review" articles to get outline
5) Filter full access papers through Pubmed Central (PMC)if no institutional access

10

10



11

11

Building the Search 2

Can perform detailed/specific search using other tools including:

1) Boolean Operators
   - Terms including "AND", "OR" and "NOT"
2) MeSH Terms
3) Clinical Queries
   - Allows narrower searching to read around clinical topics
4) Advanced Search

12

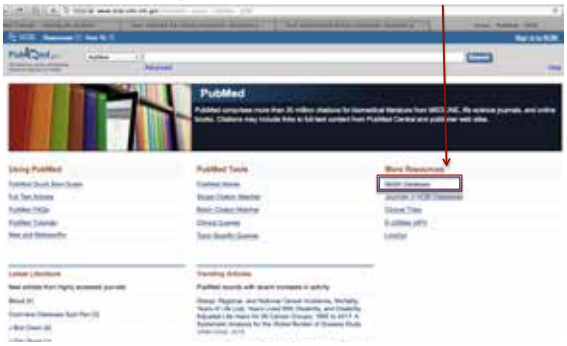
12

Example – Boolean Operators

13

13

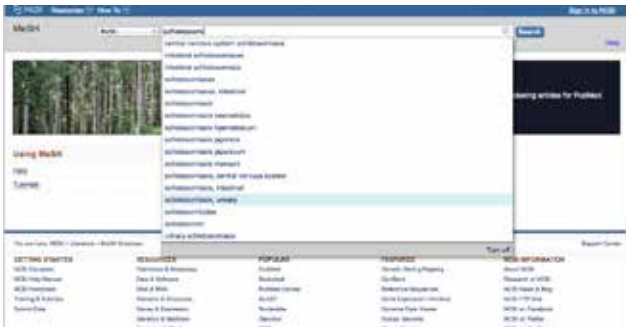
Example – MeSH Terms



14

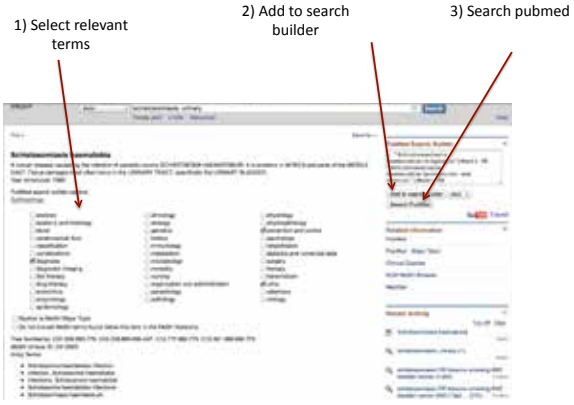
14

Example – MeSH Terms



15

15



16

16

Example – MesH Terms

17

17

Example – Clinical Queries

18

18

Manage Results

1) Too many results – refine search with filters/other methods
2) Too few results – reduce filters/use less specific terms
3) Send results to email to save for later use or save using "My NCBI" account

19

19

Hinari

- Hinari is a WHO managed site, created in 2002.
- It provides free or very low cost online access to the major journals in biomedical and related social sciences to local, not-for-profit institutions in developing countries.
- It provides information in 6 languages.

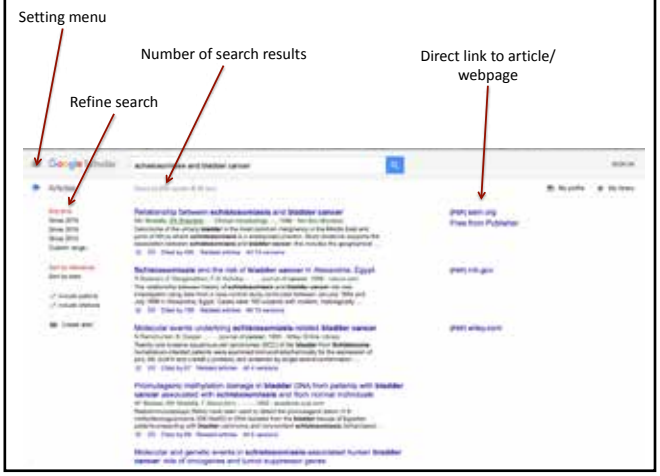
20

Google Scholar

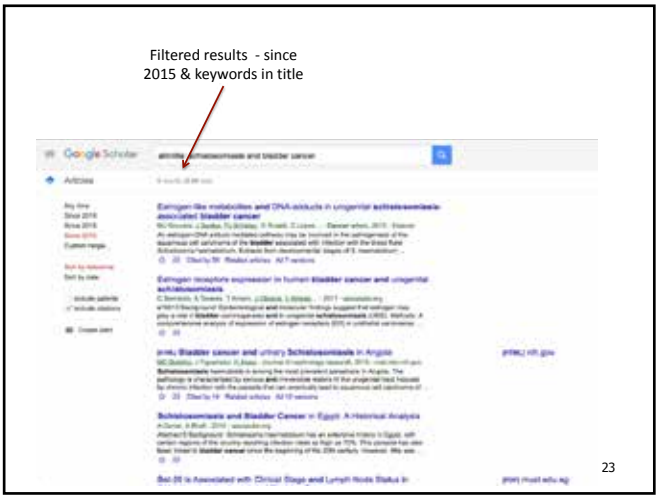
- Essentially a 'search engine' review of all 'scholarly' articles, books and other documents available on the internet
- Good: allows comprehensive searching of all online scholarly text, shows number of citations per article, easy to integrate with google account and save results
- Bad: includes more irrelevant literature, harder to use for sytematic reviews, can be harder to focus search to a few key articles

21

21



22



23

23

More Information & Help

- Pubmed
  - https://www.nlm.nih.gov/bsd/disted/pubmedtutorial
- Hinari
  - https://www.who.int/hinari/en/
- Google Scholar
  - https://scholar.google.com/intl/en/scholar/help

24

24

D2: Practical 3: Internet search engines

Using PubMed, Google Scholar and Hinari

Online practical: Go to different search engines and use your own search terms to search for articles in PubMed, Google Scholar and Hinari on topics that interest you. If you already have a good research idea or a potential research question, use this exercise to collect a list of papers you might want to read. (The next lecture will explain to you how to analyse these papers critically and to decide if they are really adding value to your background information and your own research project).

D.3: Lecture

How to Read a Research Paper Critically

Jonathan AF Hannay

1

This lecture will address:

1. Purpose of critical reading of research papers
2. Personal preparation
3. Common critical approaches
4. Types of papers and particular considerations
5. Summary
6. Acknowledgements & further reading.

- Group practice

2

1. Purpose of critical reading of research papers
2. Personal preparation
3. Common critical approaches
4. Types of papers and particular considerations
5. Summary
6. Acknowledgements & further reading.

- Group practice

3

Why bother …with critical reading?

- Preparation of self - personal learning
  - joy of learning
  - to improve your own practice and care of patients (moral duty)
- Preparation to teach
  - indirect care of patients
- Preparation for audit
  - know what the 'best practice' is
- Preparation for research…

4

- Preparation for research…
  - to know where 'the field is at'
  - to know what you don't need to do
    - saves your time & money
  - to know where the gaps / needs / conflicts / contemporary questions, etc. are.

5

"you need to read 40-50 papers before you start"
John Mendelsohn, Past President MDACC

6

6

1. Purpose of critical reading of research papers
2. Personal preparation
3. Common critical approaches
4. Types of papers and particular considerations
5. Summary
6. Acknowledgements & further reading.

- Group practice

7

Following on from the previous session on performing a literature search …you now have a large collection of papers.

Now what?…

8

Personal preparation

- Personal mind set:
  - accepting & relaxed…
  - …*vs* sceptical & inquisitive
  - …*vs* scornful.
  - critical reading identifies strengths as well as weaknesses
- Personal space:
  - quiet and well lit with dedicated time apart
- Personal tools:
  - paper pencil, highlighter,
  - ?iPad / computer for further searching
  - with a colleague?
- Personal process…

9

- Personal process…

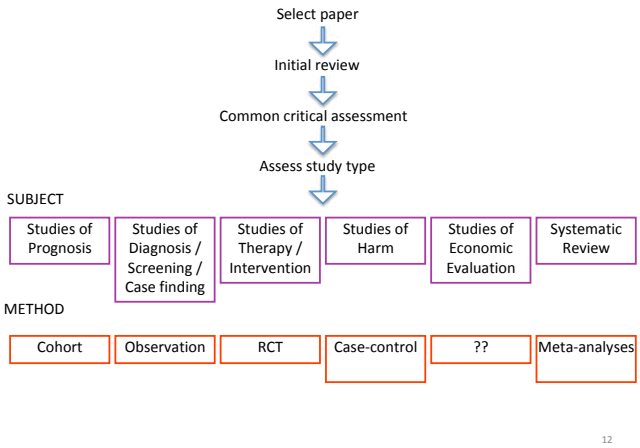
1. Read the Title and Abstract
   - (should have done this already in selection process)
2. Look at the Figures and Results
3. Read through the paper 'to get a feel'
4. Read through the paper to *scrutinise*

10

1. Purpose of critical reading of research papers
2. Personal preparation
3. Common critical approaches
4. Types of papers and particular considerations
5. Summary
6. Acknowledgements & further reading.

- Group practice

11



12

Critical approach

- There are different types of research papers but generally all follow the same framework of reporting the research:
  - Introduction / Background
  - Materials & methods / Approach
  - Results / Findings
  - Discussion / Interpretation / Conclusion

Common critical questions then subtype considerations.

13

Critiquing: Title, Authors, & Abstract

1. Search terms vs paper's true subject
   - i.e. is it really relevant to you?
     e.g. hernias in dogs vs humans; osteosarcoma vs Soft tissue sarcoma
2. Authors (and Institution)
   - Do they have appropriate expertise? (was a statistician involved??)
   - Have they reported any conflicts of interest?
     e.g. Tobacco Co. sponsorship, etc.
3. Is the Abstract informative?
   - Is an objective or hypothesis stated?
   - Are the methods appropriate for the objective?
   - Are the results complete and clear?
   - Is the conclusion sound and justified?

14


The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JUNE 30, 2011 VOL. 364 NO. 26

Mortality after Fluid Bolus in African Children with Severe Infection

Kathryn Maitland, M.B., B.S., Ph.D., Sarah Kiguli, M.B., Ch.B., M.Med., Robert O. Opoka, M.B., Ch.B., M.Med., Charles Engoru, M.B., Ch.B., M.Med., Peter Olupot-Olupot, M.B., Ch.B., Samuel O. Akech, M.B., Ch.B., Richard Nyeko, M.B., Ch.B., M.Med., George Mtove, M.D., Hugh Reyburn, M.B., B.S., Trudie Lang, Ph.D., Bernadette Brent, M.B., B.S., Jennifer A. Evans, M.B., B.S., James K. Tibenderana, M.B., Ch.B., Ph.D., Jane Crawley, M.B., B.S., M.D., Elizabeth C. Russell, M.Sc., Michael Levin, F.Med.Sci., Ph.D., Abdel G. Babiker, Ph.D., and Diana M. Gibb, M.B., Ch.B., M.D., for the FEAST Trial Group*

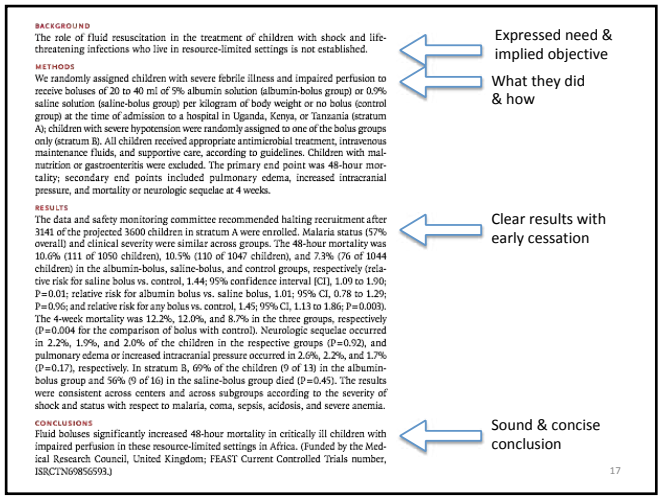
15

15

Critiquing: Title, Authors, & Abstract

1. Search terms vs paper's true subject
   - i.e. is it really relevant to you?
     e.g. hernias in dogs vs humans; osteosarcoma vs Soft tissue sarcoma
2. Authors (and Institution)
   - Do they have appropriate expertise? (was a statistician involved??)
   - Have they reported any conflicts of interest?
     e.g. Tobacco Co. sponsorship, etc.
3. Is the Abstract informative?
   - Is an objective or hypothesis stated?
   - Are the methods appropriate for the objective?
   - Are the results complete and clear?
   - Is the conclusion sound and justified?

16



BACKGROUND

The role of fluid resuscitation in the treatment of children with shock and life-threatening infections who live in resource-limited settings is not established.

METHODS

We randomly assigned children with severe febrile illness and impaired perfusion to receive boluses of 20 to 40 ml of 5% albumin solution (albumin-bolus group) or 0.9% saline solution (saline-bolus group) per kilogram of body weight or no bolus (control group) at the time of admission to a hospital in Uganda, Kenya, or Tanzania (stratum A); children with severe hypotension were randomly assigned to one of the bolus groups only (stratum B). All children received appropriate antimicrobial treatment, intravenous maintenance fluids, and supportive care, according to guidelines. Children with malnutrition or gastroenteritis were excluded. The primary end point was 48-hour mortality; secondary end points included pulmonary edema, increased intracranial pressure, and mortality or neurologic sequelae at 4 weeks.

RESULTS

The data and safety monitoring committee recommended halting recruitment after 3141 of the projected 3600 children in stratum A were enrolled. Malaria status (57% overall) and clinical severity were similar across groups. The 48-hour mortality was 10.6% (111 of 1050 children), 10.5% (110 of 1047 children), and 7.3% (76 of 1044 children) in the albumin-bolus, saline-bolus, and control groups, respectively (relative risk for saline bolus vs. control, 1.44; 95% confidence interval [CI], 1.09 to 1.90; P=0.01; relative risk for albumin bolus vs. saline bolus, 1.01; 95% CI, 0.78 to 1.29; P=0.96; and relative risk for any bolus vs. control, 1.45; 95% CI, 1.13 to 1.86; P=0.003). The 4-week mortality was 12.2%, 12.0%, and 8.7% in the three groups, respectively (P=0.004 for the comparison of bolus with control). Neurologic sequelae occurred in 2.2%, 1.9%, and 2.0% of the children in the respective groups (P=0.92), and pulmonary edema or increased intracranial pressure occurred in 2.6%, 2.2%, and 1.7% (P=0.17), respectively. In stratum B, 69% of the children (9 of 13) in the albumin-bolus group and 56% (9 of 16) in the saline-bolus group died (P=0.45). The results were consistent across centers and across subgroups according to the severity of shock and status with respect to malaria, coma, sepsis, acidosis, and severe anemia.

CONCLUSIONS

Fluid boluses significantly increased 48-hour mortality in critically ill children with impaired perfusion in these resource-limited settings in Africa. (Funded by the Medical Research Council, United Kingdom; FEAST Current Controlled Trials number, ISRCTN69856593.)

17

Expressed need & implied objective

What they did & how

Clear results with early cessation

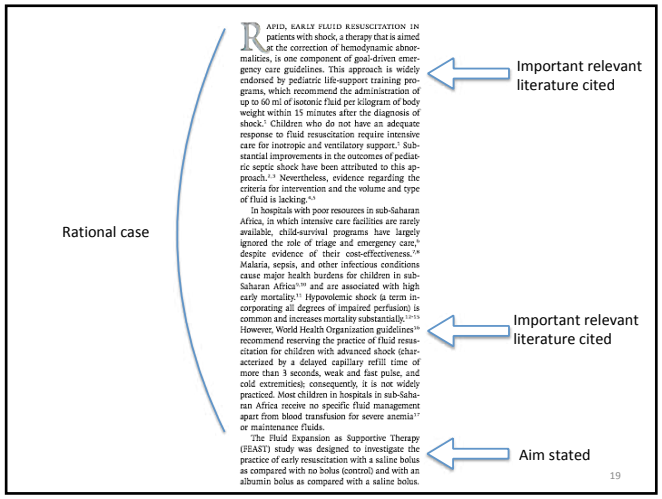
Sound & concise conclusion

17

Critiquing: Introduction and Background

1. Is an objective or hypothesis clearly stated?
   - Usually in the first or last paragraph of the section.
2. Is a rational case set out for the study?
3. Do they cite appropriate literature?
   - Are relevant important studies mentioned?
   - Are relevant important studies excluded?
   - Have they accepted discredited or dated work?
   - But remember it's not a 'review' article…
4. Is the question they seek to address relevant?
5. What type of research question is being asked?

18



Rational case

RAPID, EARLY FLUID RESUSCITATION IN patients with shock, a therapy that is aimed at the correction of hemodynamic abnormalities, is one component of goal-driven emergency care guidelines. This approach is widely endorsed by pediatric life-support training programs, which recommend the administration of up to 60 ml of isotonic fluid per kilogram of body weight within 15 minutes after the diagnosis of shock.[1] Children who do not have an adequate response to fluid resuscitation require intensive care for inotropic and ventilatory support.[2] Substantial improvements in the outcomes of pediatric septic shock have been attributed to this approach.[2,3] Nevertheless, evidence regarding the criteria for intervention and the volume and type of fluid is lacking.[4,5]

In hospitals with poor resources in sub-Saharan Africa, in which intensive care facilities are rarely available, child-survival programs have largely ignored the role of triage and emergency care,[6] despite evidence of their cost-effectiveness.[7,8] Malaria, sepsis, and other infectious conditions cause major health burdens for children in sub-Saharan Africa[9,10] and are associated with high early mortality.[11] Hypovolemic shock (a term incorporating all degrees of impaired perfusion) is common and increases mortality substantially.[12-15] However, World Health Organization guidelines[16] recommend reserving the practice of fluid resuscitation for children with advanced shock (characterized by a delayed capillary refill time of more than 3 seconds, weak and fast pulse, and cold extremities); consequently, it is not widely practiced. Most children in hospitals in sub-Saharan Africa receive no specific fluid management apart from blood transfusion for severe anemia[17] or maintenance fluids.

The Fluid Expansion as Supportive Therapy (FEAST) study was designed to investigate the practice of early resuscitation with a saline bolus as compared with no bolus (control) and with an albumin bolus as compared with a saline bolus.

19

Important relevant literature cited

Important relevant literature cited

Aim stated

19

Critical approach

Look at the Materials & Methods and Results and consider:

1. Validity
2. Importance
3. Relevance

- critical questions to assess validity, importance, and relevance will depend on the study type. e.g. therapy vs harm vs screening etc - but many questions similar.

20

Critiquing: Materials & Methods and Results

1. Ethical review / approval of the study?
2. Study design
   - is it appropriate for their research question? - Ca-Cl / Cohort / RCT…
   - If they haven't have they convincingly stated why not? - check Background section.
3. Population(s):
   - Inclusion & Exclusion criteria? and are they appropriate criteria?
   - clear definitions of what constitutes cases / conditions
4. Power
   - population sizes, length of time for occurrence of events / to see difference?
5. Outcomes
   - clear definitions of outcomes
   - subjective vs objective vs protocol assessment
   - completeness? - what about those LOST to follow-up? those who DECLINED participation?

21

6. Statistical test(s)?
   - appropriate for type of study design? for type data? - tomorrow's talk
   - was there interim analysis and did they account for this in power?
   - p-values, 95% confidence intervals
   - Odds ratio, Relative Risk, Risk reduction, NNT, NNH, Hazard Ratio,
7. Bias and confounding?
8. Conclusions from the results
   - Rational and logical
   - Do the results fit with other available evidence?
   - Satisfy Bradford Hills' criteria for causation?
     - Did exposure preceded the onset of the outcome?
     - Is there a dose-response gradient?
     - Does the association make biological sense?
     - Is the association consistent with other similar studies?
     - Is there associated evidence from a "dechallenge-rechallenge" study?
9. Are the results applicable to the local population?
   - how similar is your patient population to that of the study?
   - how similar is your local environment to that of the study?
   - are the benefit / harms / costs of the study quantifiable?

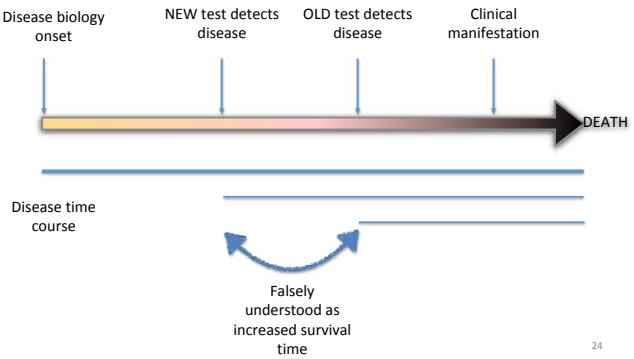
22

Considering bias

1. Pre-study bias:
   1. Weak study design - e.g. subjective assessments, lack of blinding,
   2. Selection biases
      - Lead time bias
      - Length bias
      - Channeling bias - operative choices in frail vs healthy participants
2. Intra-study bias:
   1. Interviewer bias
   2. Chronology bias - e.g. introduction of adjuncts
   3. Recall bias
   4. Attrition / Transfer bias - when losses to follow-up are uneven
   5. Misclassification bias
   6. Performance bias - e.g. inexperienced vs experienced
3. Post-study bias:
   1. Citation bias / reporting bias
   2. Confounding

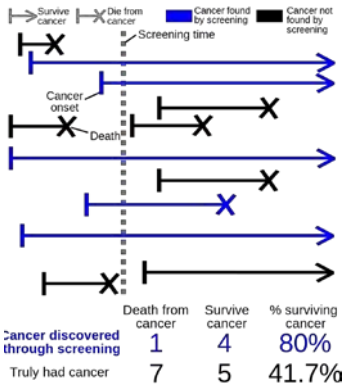
23

Lead time bias explanation



24

Length time bias



	Death from cancer	Survive cancer	% surviving cancer
Cancer discovered through screening	1	4	80%
Truly had cancer	7	5	41.7%

https://en.wikipedia.org/wiki/Length_time_bias

25

Considering bias

1. Pre-study bias:
   1. Weak study design - e.g. subjective assessments, lack of blinding,
   2. Selection biases
      - Lead time bias
      - Length bias
      - Channeling bias - operative choices in frail vs healthy participants
2. Intra-study bias:
   1. Interviewer bias
   2. Chronology bias - e.g. introduction of adjuncts
   3. Recall bias
   4. Attrition / Transfer bias - when losses to follow-up are uneven
   5. Misclassification bias
   6. Performance bias - e.g. inexperienced vs experienced
3. Post-study bias:
   1. Citation bias / reporting bias
   2. Confounding

26

Critiquing: Discussion and Conclusions

The discussion and conclusions section of a paper should:

1. Summarise the results and findings of the research,
2. Addresses potential shortcomings in the study / approach,
3. Makes clear the strengths of the study,
4. Show the relevance of the results for the research question being asked / hypothesis,
5. Explain the relationship between the results and how their answer to the research question fits into the field of knowledge,
6. Allows for relevant speculation,
7. Proposes what the next questions or required study might be.

27

1. Purpose of critical reading of research papers
2. Personal preparation
3. Common critical approaches
4. Types of papers and particular considerations
5. Summary
6. Acknowledgements & further reading.

- Group practice

28

1. Purpose of critical reading of research papers
2. Personal preparation
3. Common critical approaches
4. Types of papers and particular considerations
5. Summary
6. Acknowledgements & further reading.

- Group practice

29

Summary

1. prepare yourself well
2. you have to think FOR the authors of the papers
   1. do they understand the field?
   2. is this the right question?
   3. is this the right study?
   4. have they performed the study properly and considered potential biases and confounders?
   5. have they analysed their data properly?
   6. have they reached an appropriate conclusion?
   7. have they answered the question / hypothesis?
   8. are the criteria for validity, importance, and applicability satisfied?
3. you have to think for others
   – is this really relevant for my patients? my practice? my dept?

30

Conclusion

1. *don't waste your time*
2. *know your field*
3. *be suspicious*
4. *think **FOR** the authors of the papers*
5. *be <u>vigilant</u> for your patients:*
   1. *rejecting the influence of poor research;*
   2. *approving the adoption of good research; and*
   3. *seek to improve care with your own well conducted research*

31

1. Purpose of critical reading of research papers
2. Personal preparation
3. Common critical approaches
4. Types of papers and particular considerations
5. Summary
6. Acknowledgements & further reading.

- Group practice

32

Further information

Paper-appraisal worksheets and guides:

University of Toronto Centre for Evidence Based Medicine.
https://www.library.utoronto.ca/medicine/ebm/

CASP – Critical Appraisal Skills Programme (Oxford / NICE)
http://www.casp-uk.net/#!casp-tools-checklists/c18f8

33

D4: Practical 4

Critical reading of published papers

A number of scientific papers are made available for reading; these look good at first glance, but might have serious deficiencies. Work through a few papers, review and summarise a few; the summary must address (1) Strong points, (2) Deficiencies, (3) Will this paper change your clinical practice, and how? It works best if you discuss your ideas either in your local group or in the online chat room (one person summarises one paper but everyone reads all papers under discussion; then try to form a group opinion and test it against other groups. You will find there are no fixed rules; different readers interpret and value the same paper differently but you can come to shared conclusions by consensus. The next step is to find papers in the literature according to your own speciality or current knowledge, and analyse these critically, but always write down your opinion on points (1), (2) and (3) as above.

Please see Appendix 1 in the course handbook:

Appendix 1 contains a number of journal papers for critical reading and analysis. These papers are all available for free download through e.g. Google Scholar and copies are therefore made available in the course handbook. It works best if 1-2 persons read and summarise one paper and then present their findings to colleagues within a study group. If you give 10 minutes for each presentation and 5 minutes for discussion you can go through 5 papers in under 90 minutes. That should give you sufficient confidence to be a critical future reader.

E.1: Lecture

Collecting And Managing Your Project Data

Alison Bradley

1

Outline

- **Introduction**: *the importance of good data collection and management*
- **Types of Data**
- **Issues to Consider Prior to Data Collection**
- **Methods of Data Collection:**
  - **Primary Data Collection**
  - **Secondary Data Collection**

2

Introduction

- What is data collection?
  - "the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes"

(Kabir, 2016 in Basic Guidelines for Research: An Introductory Approach for All Disciplines Edition: First Chapter: 9. Publisher: Book Zone Publication, Chittagong-4203, Bangladesh)

3

Introduction

- What is data collection?
  - It is also a very demanding task that requires planning, organization and perseverance!

4

Introduction

- Why collect data?
  - Data collection is the most important step of the research process. Even the best research question in the world cannot be answered unless you are able to collect the required data.
  - Regardless of your field of study or whether you plan to collect data that is defined as quantitative or qualitative, accurate data collection is essential to maintaining the integrity of your research.

5

Introduction

- Why collect data?
  - The goal for all data collection:
    - capture quality evidence
    - translate this evidence into rich data analysis
    - building a convincing and credible answer to your research questions.

6

Introduction

- How is data collected?
  - 1. determine what kind of data is required
  - 2. selection of a sample from a certain population
  - 3. select an appropriate data collection instrument and clearly delineated instructions for the correct use of the instrument to reduce the likelihood of errors.
  - 4. Collect the data from the selected sample.

Each of these steps will now be looked at in more detail….

7

Types of Data

- There are generally two broad categories of data:

Qualitative and Quantitative

- The type of data you collect will depend on your research question and will also determine how you collect your data.

8

Types of Data: Qualitative

Description of Data	Mostly non numerical Descriptive Captures subjective perceptions
Types of Research Problems This Type of Data Addresses	The 'how' and 'why' of outcomes as well as effects and unintended consequences of an intervention or process To understand the process behind observed results and changes in perception
Methods of Data Collection	Data is mostly collected through focus groups, discussion groups and interviews. Unstructured methods relying on open-ended questions that fall into 3 broad categories: 1) in-depth interviews, 2)observation 3) document review. Protocol is less structured so the researcher may change their data collection strategy by adding or refining their informants. Triangulation is used to increase reliability of findings which means that the researcher often uses multiple methods of data collection. The researcher needs to record all useful data thoroughly, accurately and systematically often as field notes, audio or video tapes

9

Types of Data: Qualitative

Advantages	Can improve qualitative survey-based methods by strengthening design and generating evaluation of hypothesis, expanding or clarifying findings
Disadvantages	Expensive Time consuming Findings non-generalizable

10

Types of Data: Quantitative

Description of Data	Numerical
Types of Research Problems This Type of Data Addresses	Addresses the 'what' of a problem
Methods of Data Collection	Systematic standardized approach using structured data collection and instruments to observe and record well defined events Mostly measures something using different standardized scales (for example nominal, ordinal or ratio scales) Random sampling Often data is obtained from management information systems, surveys with closed questions and structured interviews

11

Types of Data: Quantitative

Advantages	Cheaper Results are more comparable and generalizable Size of effect can be measured
Disadvantages	Limited capacity to explain or investigate unexpected similarities or differences

12

Types of Data: Mixed Methods

- Sometimes qualitative and quantitative research design, data, techniques and methods can be combined
- This means a number of different methods for data collection can be used within a study to capitalize the strengths and minimize the weaknesses of taking a single approach

13

Types of Data: Mixed Methods

- Mixed methods are useful to address complex questions of research addressing marginalized populations
- Mixed methods are often used to: initiate, design, develop and expand interventions, perform evaluation, improve research design and corroborate or triangulate findings
- Challenges for this approach include: delineating complementary research questions and selecting which research methods to combine. Data collection and analysis can also be time consuming.

14

Types of Data

- Regardless of whether using qualitative, quantitative or mixed methods the data collected can be primary data or secondary data.

15

Types of Data: Primary Data

- Primary data is collected first-hand
- It is considered to be superior to secondary data as it has not been altered therefore its validity is greater than secondary data
- It is more reliable and objective
- Primary data has not yet been published

16

Types of Data: Primary Data

- Sources of Primary Data:
  - Experiments
  - Survey
  - Questionnaire
  - Interview
  - Observations

17

Types of Data: Primary Data

Advantages of Using Primary Data	Disadvantage of Using Primary Data
• **You collect data specific to the research problem.** • **You can defend the quality of the data you collected** • **Additional data can be collected if required**	• **Time consuming:** deciding why, what, how and when to collect data; Collecting data(or coordinating others to do so); Dealing with ethical issues (informed consent, data protection) • **Ensuring high standards of data collection are maintained:** Ensure that data is obtained accurately, in the correct format; no fake data, unnecessary data not included • **Cost**: including acquiring funding

18

Types of Data: Secondary Data

- Secondary data comes from a source that has already been published (e.g. a literature review)
- It is data that has been collected by someone else for another purpose but is being used by the researcher for their own purpose
- Although secondary data is less valid it can still be important for example when:
  - It is difficult to obtain primary data or primary data simply does not exist
  - Primary data exists but respondents are unwilling to reveal it

19

Types of Data: Secondary Data

- Sources of secondary data include:
  - Books
  - Census data
  - Records
  - Biographies
  - Newspapers
  - Data archives
  - Internet articles
  - Research articles
  - Databases etc.

20

Types of Data: Secondary Data

Advantages of Using Secondary Data	Disadvantages of Using Secondary Data
• Less expensive • Secondary data generally has a pre-established degree of validity and reliability • Can be helpful in the research design of subsequent primary research • Can provide a baseline against which primary data results can be compared • It is always advisable to begin any research with a review of the secondary data.	• Reliability and accuracy of data can be questionable • Data collected in one location may not be applicable to another location • Data may become outdated • Secondary data may need to be amend or modified for use. • Potential authenticity and copyright issues

21

Recap…

- By now we have established
  - The importance of the data collection stage of the research process
  - The different types of data collected for both qualitative and quantitative research
  - The advantages and disadvantages of using primary and secondary data for either qualitative or quantitative research

  Now we will look at ethical considerations that need to be considered prior to collecting data before looking at how to go about collecting primary and secondary data

22

Ethical Issues to Consider Prior to Collecting Data

Many professional, educational and government institutions and bodies have adopted specific ethical codes and policies.

- Ethical norms in research are important for the following reasons:

1. They promote knowledge, truth, and avoidance of error as the aims of research by avoiding fabricating or misrepresenting data
2. They facilitate collaborative working by promoting trust, accountability, respect, and fairness. This is evident in guidelines for authorship and data sharing, copyright and patenting, and blind peer review.
3. They ensure researchers are held accountable to the public.
4. By increasing trust in quality and integrity of research they help to build public support and funding for research
5. They promote moral and social values: for example human rights and animal welfare

23

Ethical Issues to Consider Prior to Collecting Data

- Honesty: in all aspects of the research process including reporting methods and results of research
- Objectivity: avoid bias
- Carefulness and Competence: avoid careless errors and negligence, commit to life long learning
- Openness: data sharing
- Intellectual Property: patents, copyrights etc. must be respected and plagiarism avoided
- Confidentiality
- Mentoring and Respect for Colleagues
- Social Responsibility
- Legality: obey relevant laws and policies

24

Methods of Data Collection: Primary Data

There are several methods of collecting primary data including but not limited to:

- Questionnaires/Surveys
- Interviews
- Observation
- Experiments
- Others: Case-studies, Focus Group Interviews, Diaries, Process Analysis, Statistical Method etc.
- We will now focus on some of the more common methods: questionnaires, interviews, observation and experiments

25

Methods of Data Collection: Primary Data: Questionnaires

Types of Questionnaire	Measure separate variables OR aggregate into a scale or index	
Types of Questions	**Closed Questions**	**Open Questions**
	Options must be exhaustive and mutually exclusive Can be: • Dichotomous (2options) • Nominal-polytomous (2+ unordered options) • Ordinal-polytomous (2+ordered options)	Must be sequenced: least to most sensitive, factual to attitudinal, general to specific Sequence of question types: Screening: gauge whether the person should complete the questionnaire Warm-up: easy to answer, captures interest Transitional: connects sections of the questionnaire Skips: for example "if no go to question 6" Difficult questions: near the end Changing formulae: Demographic questions at the end
Item Construction	Clear wording with terms understood to have the same meaning across different subgroups of the study population Include an 'open' answer box after a list of options 1 question per item ensuring different options reflect different opinions Use positive statements No leading or biased questions	

26

Methods of Data Collection: Primary Data: Questionnaires

Administration	Face-to-face Post Computer
Advantage	Large amount of data from large population in short period of time Cost effective Easily quantified Objectively analysed Results can compare/contrast/measure change Can be repeated at set time interval
Disadvantage	No adequate for all types of information (e.g. feelings/emotions) Limited to scope of questions posed Truthfulness of respondents not guaranteed Questions open to interpretation so always will have an element of subjectivity Timing of questionnaire could bias results

27

Methods of Data Collection: Primary Data: Interviews

- Interviews can be
  - Structured
  - Semi-structures
  - Unstructured

28

Methods of Data Collection: Primary Data: Structured Interviews

Characteristics	When to Use	How to Record	Benefits
• Same series of questions asked • Questions are created prior to the interview • Few open-ended questions • Limited response categories • Ordering and phrasing of the questions are consistent • Interviewer plays a neutral	• Well-developed understanding of the topic. • Focused research question	Paper-based recording Face-to-face interview Audio recording Video recording Telephone interview Web-based interview/recording Self-reporting	Efficient Comparable results Do not require extensive interviewer training

29

Methods of Data Collection: Primary Data: Semi-Structured Interviews

Characteristics	When to Use	How to Record	Benefits
• A formal interview. • A list of questions and topics are formulated to guide the interview but the interviewer is able to add topical trajectories	• If you won't get more than one chance to interview someone again • You will be sending several interviewers to collect data. • Understanding of the topic already developed through observation, informal and unstructured interviews	Paper-based (have a note taker) Audio or recording that is later transcribed	• Questions can be prepared ahead of time. • Allow informants the freedom to express their views in their own terms. • Provide reliable, comparable qualitative data • Provide the opportunity for identifying new ways of seeing and understanding the topic at hand.

30

Methods of Data Collection: Primary Data: Unstructured Interviews

Characteristics	When to Use	How to Record	Benefits
Interviewer builds rapport with respondents who express themselves in their own way to open-ended questions .	• Researching an 'as-of-yet' poorly understood topic. • Interviews can take place multiple times	Audio or video recordings later transcribed	• Can test researcher's preliminary understanding • New ways of understanding may develop

31

Methods of Data Collection: Primary Data: Observation

- There are several types of observational methods (naturalistic, structured, unstructured, participant, non-participant)
- How the results of observational research are analyzed and reported depends on how they are recorded
- There are several approaches:
  - Field notes: (creating a template to guide observations, observational coding sheet)
  - Qualitative records(detailed observations reordered without predetermined categories or questions)
  - Quantitative measures (frequency, duration of occurrence

32

Methods of Data Collection: Primary Data: Experimental

There are three types of experiments:

- 1. Laboratory / Controlled Experiments
- 2. Field Experiments: independent variables are still manipulated but in a real-life setting.
- 3. Natural Experiments: conducted in real-life with no control over events

33

Methods of Data Collection: Secondary Data

- Secondary data can be obtained from two different research strands:
  - Quantitative: Census, housing, social security and other databases
  - Qualitative: interviews and focus groups transcripts, field notes, observation records etc.

And can be found in multiple sources including but not limited to:

- Published Printed Sources
- Books
- Journals
- Newspapers
- Published Electronic Sources
- e-journals
- Websites
- Unpublished Personal Records: e.g. diaries, letters
- Government Records
- Census Data

34

Conclusion

- Data collection is an important step in the research process and requires significant planning and time allocation
- The type of data you collect and how you collect it will depend on your research question and whether you are undertaking qualitative or quantitative research
- When planning data collection you must consider:
  - What type of data you need to in order to answer your research question
  - How much data (sample size) do you need to answer your research question
  - How you can obtain/access this data and any ethical issues you need to consider, how much time and resources you can dedicate to accessing this data
  - How you will collect and store your collected data and any ethical issues you need to consider (confidentiality, data protection laws and policies etc.), what resources (time, money and human resources) you can dedicate to this task
- Remember data collection is an essential step in the research process and is a very demanding task that requires good planning, organization and perseverance!

35

E2: Practical 5

Writing a detailed study proposal

The aim is that you and/or your group will now be able to write down a study proposal with appropriate research title, research question, study design, planned methods. Also discuss specific practical problems in research methods that you have encountered or envisage.

Start to ask yourself:

- What is your research question? Is it robust and unambiguous?
- How to select the correct study design for your research question
- What research method(s) will work best for your question?
- End points, variables, feasibility?
- How will data be collected and who will do it?
- What is the desired and/or available time frame to complete data collection?
- How will data be analysed and by whom?
- How to prepare for ethical approval for your study.
- Time management.
- Recruiting other team members.
- Reviewing end-points regularly.

Medical Statistics

"The Magical Mystery Tour"

Noel Thin
Vanessa Fawcett

COSECSA
College of Surgeons of East, Central and Southern Africa

1

For many doctors the thought of medical statistics can be scary.....

2

Or confusing.....

3

Hopefully we can simplify it !

4

Summary

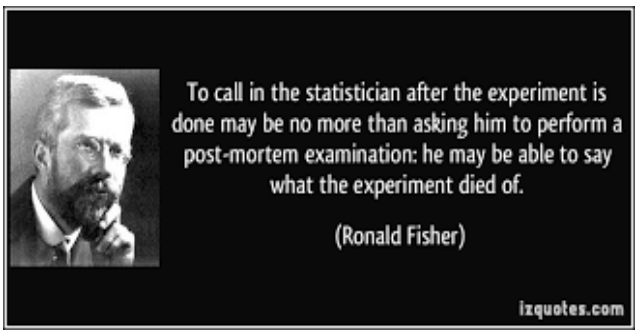
- Understanding basic statistical concepts is central to understanding the medical literature
  - And to doing research yourself!
- It is **NOT** important to understand the tests or the underlying math
- **<u>You need to know when a test should be used and how to interpret its results</u>**

5

Words of Encouragement

- You are not alone
- Get access to a good basic statistics book, e.g.
  - Medical Statistics at a Glance – Aviva Petrie & Caroline Sabin
  - Biostatistics for Dummies
- Knowing statistics does not make you clever but it does make you useful
- Make friends with a statistician from the start!
  - Usually the grumpy one, drinking coffee, in the corner of the room, staring intensely at the computer

6



7

Aims of this lecture are to understand:

- Principles of descriptive statistics
- Principles of inferential statistics
- Data and hypothesis testing
- Predictive modelling

8

An example of a study …

Post-discharge functional outcomes of trauma patients:

- All injured patients admitted to hospital
  - Standardized telephone interviews
  - 1 and 3 months post-discharge
- Outcomes based on a standardized scale
  - 1=Death
  - 8=Excellent recovery

9

Ask yourself…

- Who is the population?
- What type of study is this?
- What are some of the data points that we are going to be collected?

Think about these as we move forward.

10

What is Data?

- Data is obtained from a sample which represents the population

- We are usually using data to look for an **effect**
- Statistics: the method of collecting, summarizing, analysing & drawing conclusions from data

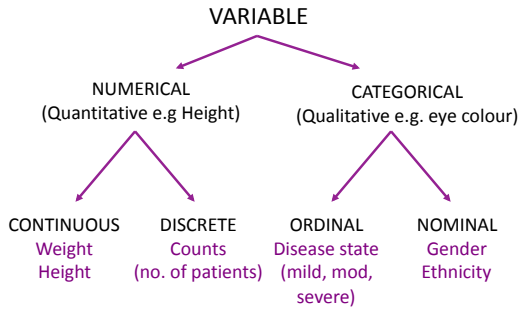
11

The distinction between sample and population is key…

- Example:
- If I want to know about the favourite football team supported by people in Zambia, the people of Zambia are the population, but we then choose a smaller proportion, the sample, in order to make generalizations about the population. The people in a room could be an example of a sample.

12

Types of Data-1: Measured/Collected



13

Types of data-2: Derived

DERIVED DATA

Means there is a Process of formatting the collected data to produce:

- Percentages/Proportions
- Ratios
- Rates
- Scores

14

Types of Data - Why important?

- The type of data defines:
  - The summary measures used for analysis (how to describe your data):
    - Mean (+Standard deviation), Median (+Range) for continuous data
    - Proportions for discrete data
  - The statistics used for analysis

15

Types of Statistics/Analyses

A. Descriptive Statistics
- Frequencies
- Basic measurements

Describing a phenomenon
How many? How much?
BP, HR, BMI, IQ, etc.

B. Inferential Statistics
- Hypothesis Testing
- Correlation
- Confidence Intervals
- Significance Testing
- Prediction

Drawing conclusions
Proving or disproving theories
Associations between phenomena
If sample relates to the larger population
e.g. Diet and health

16

A. DESCRIPTIVE STATISTICS

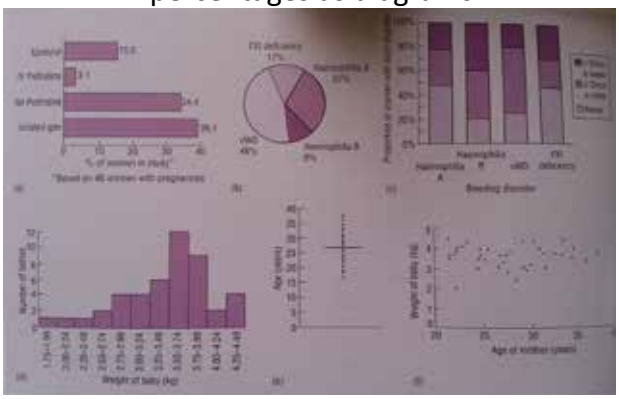
Descriptive statistics can be used to summarise and describe variables

"What does the data look like?"

- Frequencies (counts) & Percentages
- Central tendencies ("averages") & Spread
  - Use with numerical data
    - Height, weight, cholesterol, scores on a test

17

Helpful to display frequencies & percentages as diagrams



18

Diagrams and Graphs

- Can all be used to display frequencies or counts.
- Gives you a real sense of what the data looks like.
- Can be used for numerical or categorical data.
- Is a good starting point to show what results you have after you have collected data.

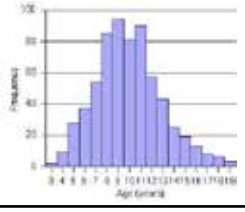
19

Histogram

Continuous or Discrete Data

- One of the most common ways to display frequencies of data
- Used for discrete or continuous data
- Shows distribution of a continuous data set

e.g. Ages of individuals in a sample



20

Numeric Descriptive Statistics

1. Measures of central tendency ("averages") of data:
   Mean, Median, Mode

Because there is no certainty in life (or statistics) we need to know the accuracy i.e. through…

2. Measures of spread (variability) of data
   - Standard Deviation
   - Inter-quartile range

21

- Remember that this is for numeric data, e.g:
- For the study I'm doing, I want to know the average score between 1 and 8; that is the central tendency.
- If almost everyone has scores between 6 and 8, they are all doing very well, then my variability will be very low.
- But if the scores go all the way from 1 and 8 there may be high variability.

22

Numeric Descriptive Statistics: Central tendencies

MEAN	**Average or arithmetic mean of the data**
MEDIAN	**The value which comes half way when the data are ranked in order**
MODE	**Most common value observed**

- In a normal distribution, mean and median are the same
- The mode is of little if any practical use

23

Mean

- Definition:
  - Sum of all the values in a sample, divided by the number of values
- Best applied in normally distributed, continuous data
  = **Normal (Gaussian/Parametric) distribution**

24

Histogram

Numeric Data

- E.g. heights of individuals in class

No. of children at that height

Height of children

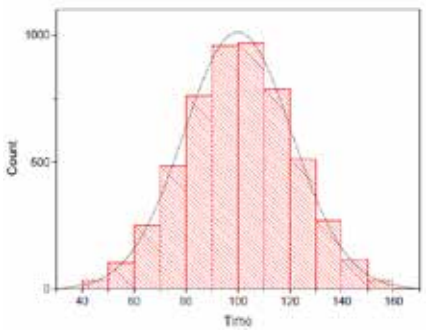
25

Notes on next slide

- If we then take the histogram and draw a curved line that fits the outline of the bars we can then describe the shape of the line.
- If the line is symmetric like this, with one peak in the middle, then the distribution is considered a normal distribution.

26

Histogram



This is a symmetrical "Bell Shaped" curve of a Normal Distribution

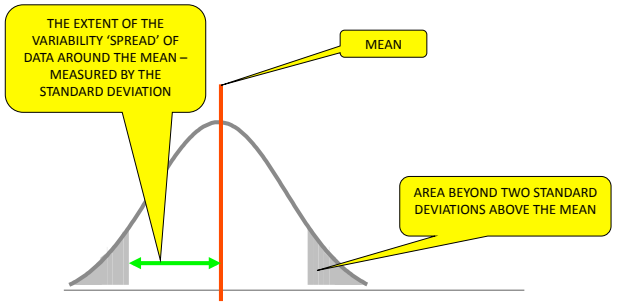
27

Notes for next 2 slides

- Looking at the curve of a continuous data set, we can describe the main features that descriptive statistics describe:
  - central tendency
  - variability
- Standard deviation really describes the Variability

28

Normal Distribution

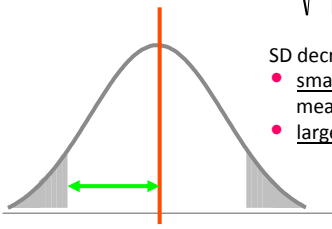


29

Standard Deviation

is a measure of the Spread of Values of the Sample Around the Mean

$$SD = \sqrt{\frac{\text{Sum(Value – Mean)}^2}{\text{Number of values}}}$$

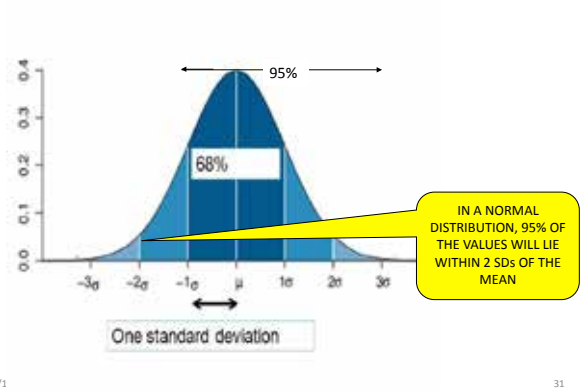


SD decreases as a function of:

- smaller spread of values about the mean
- larger number of values

30

Standard Deviation (σ)



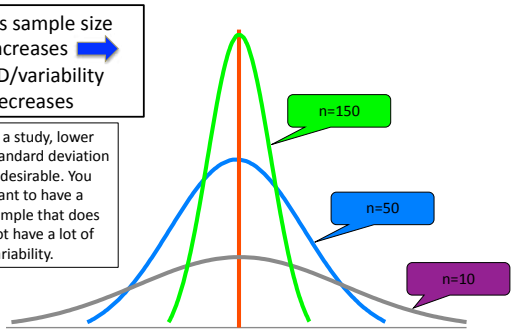
9/1 31

31

Standard Deviation & Sample Size

As sample size increases ➡ SD/variability decreases

In a study, lower standard deviation is desirable. You want to have a sample that does not have a lot of variability.



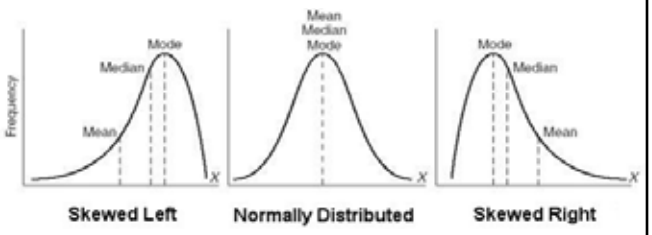
32

Why this is important

- If sample size is large enough, most data sets will display normal distribution
- This distribution underpins all statistical probability testing
- This principle is used to test a hypothesis
- and to quantify possible statistical errors

33

Skewed Distribution



- But not all distributions will appear normal, some are skewed.
- With skewed distributions the Mean is not a reliable reflection of central tendency.
- With skewed distribution we therefore use other descriptive and inferential statistics.

34

Median

- Definition:
  - Middle value from an ordered listing of the values
    - If an odd number of values, it is the middle value
    - If even number of values, it is the average of the two middle values
- Better to indicate the "average" in a non-normal distribution

1	7	7	8	10	15	21	35	43

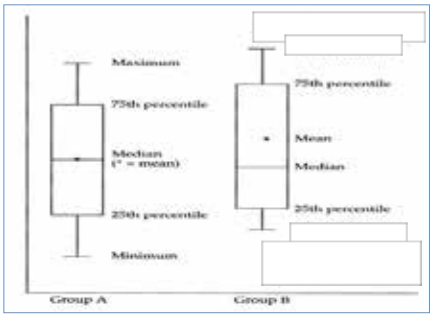
35

Quartiles

- Description of spread in non-normal data set
- e.g. 25th percentile: 25% of the values are below this level, 75% of the values are above
- Interquartile range (IQR)
  - Is the range of data from the 25th percentile to the 75th percentile
  - The midpoint is the 50th percentile = median

36

Box and Whiskers Plots



See next slide for explanation

37

About Box and Whiskers plots

- So in the same way that we used a histogram and the normal curve to show a picture of what normal distribution of data looks like, this is a way to show the central tendency and variability of non-normal distribution.
- It is thus a pictorial way to show a non-parametric data set with full range of values, the inter-quartile range and median.
- The box and whiskers on the left is actually a normal distribution, shown a different way.
- The one on the right is a non-normal distribution. The spots that are far out are all data points that are outliers. Think if you had a trainee in your group that was 78 years old. If you tried to calculate a mean it would seem much higher than the average of the group.
- So the mean is sensitive to outliers, but median not so much.

38

The type of distribution affects choice of both descriptive and inferential statistical tests

- Normal distributions = Parametric data sets
  - Described using Mean & Standard Deviation
  - Use *parametric tests*
- Non-normal distributions = Non-parametric data sets
  - Describe using median and Inter-Quartile Range
  - Use *non-parametric tests*

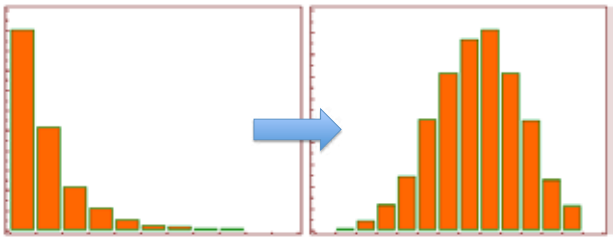
39

Transformation of data

- Sometimes, a non-normal distribution can be "transformed" to a normal distribution
  - Then you can use parametric descriptions and statistical tests (generally more accurate!)
- Can mathematically transform by taking square root, calculating logarithm etc.
- Must remember to transform data results back to original units and clinically correlate

40

Transformation of data



transformed by taking
the log of the values

41

The type of data will dictate the type of Statistics/Analysis you will use

A. Descriptive Statistics
- Frequencies
- Basic measurements

Describing phenomena
- How many?
- How much?
- BP, HR, BMI, IQ, etc.

B. Inferential Statistics
- Hypothesis Testing
- Correlation
- Confidence Intervals
- Significance Testing
- Prediction

Drawing conclusions
- Proving or disproving theories
- Associations between phenomena
- If sample relates to the larger population
- E.g., Diet and health

42

B. PRINCIPLES OF INFERENTIAL STATISTICS

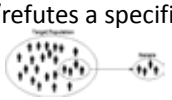
To compare samples and make inferences (or draw conclusions) regarding the population itself:

i. Hypothesis testing
ii. P values
iii. Sampling distribution + Standard Error (SE)
iv. Confidence intervals
v. Statistical Errors
vi. Effect of sample size (power calculation)

43

(i). To compare effects you need a **hypothesis**

- We gather sample data to assess how much evidence we have that supports/refutes a specific hypothesis about a population



- Does our data represent real relationships or random fluctuations (by chance)?
- Use hypothesis testing to quantify the sample data against the true state of the population

44

Extra notes on Hypothesis

- In order to compare our collected variable to either another population or another variable, we need to begin with a hypothesis.
- The hypothesis may be about our population of interest compared to another population of interest.
- Or it may be about a characteristic/variable in our population, and how it relates to another variable in our population.
- Example: How does the average age of surgical trainees in Zambia compare to the average age of obstetric trainees? Maybe our hypothesis is that surgical trainees are older.
- Example: How does the average age of surgical trainees relate to the likelihood of being married? Maybe our hypothesis is that older trainees are more likely to be married (Or maybe it's the other way round!).

45

Hypothesis Testing: What are the steps?

1. Define the null and alternate hypotheses.
2. Collect relevant data from a sample of individuals.
3. Use an appropriate test for the sample data and calculate a test statistic
   (through a Formula or Computer programme)
4. Interpret the P-value and result.

46

The Hypothesis has to be a Precise Statement

- **Null hypothesis = H_0 :** There is no effect/variation between variables of interest in the population
- **Alternate hypothesis = H_1 :** There is an effect/ variation between variables of interest in the population
  - AND this effect is likely **not** due to chance.
- We can only accept H_1 if we statistically **reject the null hypothesis**
- Please note we do not say that we accept the alternate hypothesis but that we reject the Null hypothesis. We can only accept or reject the Null hypothesis.

47

Example

- Post-discharge functional outcomes of younger versus older trauma patients
  - Who has better outcomes after trauma, older or younger patients?
- **Null hypothesis = H_0** : Younger patients have similar outcomes to older patients after trauma
- **Alternate hypothesis = H_1** : Younger patients have better outcomes after trauma compared to older patients

48

Example (page 2)

- Collect data
  - Telephone interviews
  - Calculate "outcome after trauma" scores
- Compare the mean scores between older and younger patients
- H_0: $Mean_Y - Mean_O = 0$ = No difference!
- H_1: $Mean_Y - Mean_O > 0$ = Younger do better!

49

```
. ttest GOSE_Y == GOSE_O, unpaired

Two-sample t test with equal variances

Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   GOSE_Y |      40        5.45    .3507318    2.218223    4.740578    6.159422
   GOSE_O |      40       3.775    .3694894    2.336856    3.027637    4.522363
---------+--------------------------------------------------------------------
combined |      80      4.6125    .270076    2.415633    4.074927    5.150073
---------+--------------------------------------------------------------------
    diff |            1.675    .509446                .6607709    2.689229
------------------------------------------------------------------------------
    diff = mean(GOSE_Y) - mean(GOSE_O)                            t =  3.2879
Ho: diff = 0                                     degrees of freedom =      78

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9992         Pr(|T| > |t|) = 0.0015          Pr(T > t) = 0.0008
```

We can see from the calculation that there is a difference in the mean scores between younger and older but is this statistically strong enough to reject the null hypothesis. **For this we need to calculate the P-value**.

50

(ii). P values

- The probability that we find an effect in our study when there is not one in the population (i.e. the finding is due to chance)
- Traditionally:
  - If P>0.05 we accept the null hypothesis
  - If P≤0.05 then we have enough evidence to reject the null hypothesis
- But:
  - The P value used is an arbitrary value!
    - P value of 0.05 equals 1 in 20 (5%) chance
    - P value of 0.01 equals 1 in 100 chance
    - P value of 0.001 equals 1 in 1000 chance

You can set this

51

P values summary

- P value≤0.05
  Reject the null hypothesis and say the results are statistically significant at the 5% level
- P>0.05 there is not enough evidence to reject the null hypothesis (does not mean null hypothesis is true)
  We do not reject the null and say the results are not statistically significant at the 5% level

52

(iii). Standard Error

- We need to look at the accuracy of the actual result… does it make sense?
- Standard error is a measure of the precision of a sample in estimating the population parameter…
  - Precision reflects random fluctuations in data
- **Important:** SE depends on sample size (Similar to SD)
  - The larger the sample, the smaller the standard error.

Standard error of the mean (SEM)
=Standard deviation / square root of (sample size)

Standard error of the proportion
=Square root of (proportion X 1 - proportion) / sample size)

53

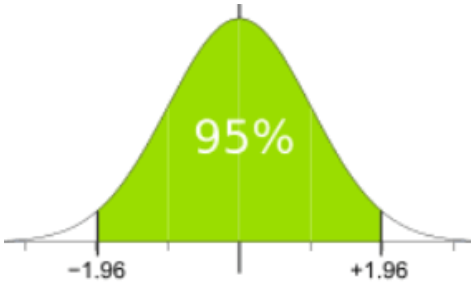
(iv). Confidence Intervals

- A range of values that is likely to contain the "true" population value
- Based on SE, but CIs are more meaningful clinically
  - Also are based on the "population", not the "sample"
- Most common CI = 95% CI

A 95% CI is defined by:
Sample mean (or proportion) ± 1.96 X standard error

54

Confidence Intervals



55

Confidence Intervals

```
. ttest GOSE_Y == GOSE_0, unpaired

Two-sample t test with equal variances

Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   GOSE_Y |      40        5.45    .3507318    2.218223    4.740578    6.159422
   GOSE_0 |      40       3.775    .3694894    2.336856    3.027637    4.522363
---------+--------------------------------------------------------------------
combined |      80      4.6125    .270076    2.415633    4.074927    5.150073
---------+--------------------------------------------------------------------
    diff |            1.675    .509446                .6607709    2.689229
------------------------------------------------------------------------------
    diff = mean(GOSE_Y) - mean(GOSE_0)                            t =   3.2879
Ho: diff = 0                                     degrees of freedom =       78

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9992         Pr(|T| > |t|) = 0.0015          Pr(T > t) = 0.0008
```

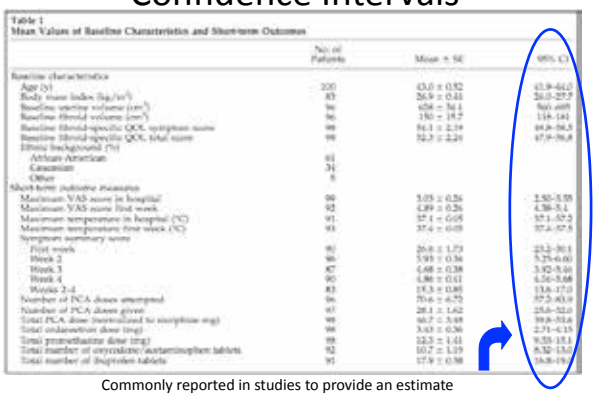
56

Some explanation

- Precision (measured as CIs) reflects random fluctuations in the data, and how well our sample is representing our population of interest.
- So you want the CI to be narrower, which means your sample is more likely to closely represent the population .
- You may also find specific clinical value in the high and low values, or values that are in the range.

57

Confidence Intervals



Commonly reported in studies to provide an estimate of the precision of the mean.

58

What does it mean?

- We are 95% confident that the <u>true</u> mean of the <u>population</u> lies between the two points.
- See how wide the CI is for feel of precision (narrower is better).
- Do these upper and lower figures have clinical significance?
- Does the range include values of particular interest?

59

P values and Confidence Intervals

- P values provide less information than confidence intervals.
- P value provides only a probability that estimate is due to chance.
- P value could be statistically significant but of <u>limited clinical significance</u> e.g. a study with very large numbers might find that a difference of 0.1 which is statistically significant but on a scale of 0-10 it may be of no clinical significance.
- P values tell us more about our study, and sometimes the size of our sample, rather than our actual population.

60

(v). What About Error?

- You have collected the data,
- Done the statistics,
- Have a P value <0.05 and acceptable SD and CI,
- So a statistically significant difference;
- You have rejected the null hypothesis:
- Job done – congratulations - pat on the back!!
- Not quite – is this a true result in the population?
  (remember that you have only studied a sample)

- **Now ASK:**
  - Is your experiment adequately **powered** to give a true result?
  - Are your samples **big enough** to detect a true result?

61

Statistical Errors

A. Type I error:

- Rejects the null hypothesis when it is true
- Claiming an effect when in fact there is none
- Also called the α error
- Typically 0.05 is used
- This is the significance level of test i.e. P<0.05
- So when 0.05 is used, you are accepting that there is a 5% chance of making the error that your results showed a difference but there is actually no difference.

62

Statistical Errors

B. Type II error

- Fails to reject the null hypothesis when it is false
- Claiming there is no effect when in fact there is.
- Also called a β error. Usually 20% (0.20)
- The probability of not making a Type II error is 1 - β, which is called the **power** of the test (80% or 0.80)
- Power calculation (sample size calculation): make sure your study has a large enough sample to give adequate power. If sample is not big enough it is a hidden error.

Otherwise you may not find a difference in compared samples when there really is one in the population

63

Statistical Errors: Summary

Type I (α)	• **Find a significant difference even though one does not exist** • **'False positive'** • **Usually set at 0.05 (5%)**
Type II (β)	• **Fail to find a significant difference even though one exists** • **'False negative'** • **Usually set at 0.20 (20%)** • **Power = 1 – β (i.e. usually 80%)**

Remember that power is related to sample size because a larger sample has a smaller SE

64

Statistical Errors

Population Reality

Test Result		Null Hypothesis false	Null Hypothesis true
	Null Hypothesis reject	No Error	Type I α
	Null Hypothesis is not rejected	Type II β	No Error

65

(vi). Sample Size Calculation

- Also called "power analysis"
- When designing a study, one needs to determine how large a study is needed to maximize the ability to detect the truth.
- Power is the ability of a study to avoid a Type II error
- Power = (1-β)
- Goal: Maximize ability to find the population "truth", while minimizing risk of errors
  - Many studies are completed without proper estimate of appropriate study size.
  - This may lead to a misleading study outcome

66

Sample Size Calculation

- Information needed to calculate:
  - Level of Type I error: 0.05 typical
  - Level of Type II error: 0.20 typical
  - Power = 0.8
    - It is ethically irresponsible + wasteful of resources to plan with a higher risk of Type II error
  - **Need to have an idea of Inherent variability of population (mean and SD)**
    - *****Usually estimated from preliminary data (prior similar studies or pilot data)**
  - Null hypothesis
- There are different formulas you can use to estimate sample size, depending on type of study.
- Statistical software will have programmess that can calculate this.

67

Sample Size Calculation: example

```
. sampsi 20 30, alpha(.05) power (.8) sd (10)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                  and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            m1 =       20
            m2 =       30
           sd1 =       10
           sd2 =       10
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =       16
            n2 =       16
```

Stata Input: Mean 1 = 20, mean 2 = 30, α = .05, power (1-β) =.8, std. dev. 10.
To do the study without error we need at least 16 patients in each group

68

What does this mean?

- The sample size calculation is the minimum number of individuals required in each group to <u>prove a true effect, if it exists</u>
- Pragmatism

- Although theoretically possible, a properly powered study may be too big to recruit for…
  - Time, money, resources available; so…
  - You may need to do a pilot study to find
    1. estimates of means & SDs for your power calculation
    2. feasibility of undertaking the study

69

STATISTICAL TESTS

<u>(You do not need to know these in detail, so this is just a brief summary of tests available)</u>

- Type of statistical test used is influenced by:
  - Type of data (numerical vs. categorical)
  - Number of groups
  - Whether sample data is normally distributed
  - Whether individuals in the treatment group are paired with individuals in the comparator group
- Try to use the most appropriate test for the data to improve the accuracy of the results

70

http://www.ats.ucla.edu/stat/mult_pkg/whatstat/

71

Numerical data: a single group

Comparison of known mean to mean of your sample

- <u>Parametric test</u>
- Large sample size
- Normal distribution
- Means + SD

- If we do not know population variance – use one sample t-test
- If we know variance – use z-test (almost identical to t-test)

- <u>Non-parametric</u>
- Skew or small sample size
- Medians

- Use Sign test

72

Numerical data: a single group

Comparison of known mean value to mean of our sample

- **Example: Interested in whether the systolic blood pressure of our sample of diabetics is the same as the population SBP (120)**
- **H_0=Our sample mean is not different from the population mean**

- P=0.03 therefore we can reject the null hypothesis
- Result: Strong evidence to reject the null hypothesis

73

Numerical data: two related groups

Study: Comparison of 2 samples related to each other
e.g. same individual variable at different time points
e.g. individually matched – case control study

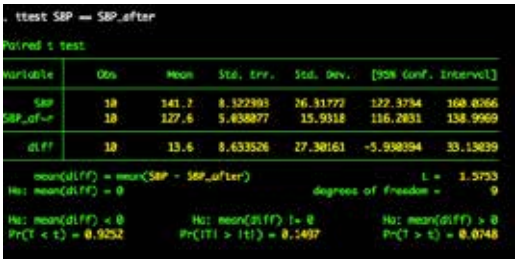
- **<u>Parametric data</u>**
- same sample size in each group
- Normal distribution
- Means and SD
- **Use paired t-test**

- **<u>Non-parametric data</u>**
- Skew or small sample size
- Medians
- **Use Sign test or Wilcoxon signed rank test**

74

Numerical data: two related groups

- **Example: Now we want to compare the SBP of our sample before and after a new BP medication**



- P=0.15 therefore we cannot reject the null hypothesis

75

Numerical data: two unrelated groups

Study: Comparison of 2 samples unrelated to each other

- **<u>Parametric data</u>**
- Normal distribution
- Means and SD
- **Use unpaired (two-sample) t-test**

- **<u>Non-parametric data</u>**
- Skew or small sample size
- Medians
- **Use Wilcoxon rank sum (two-sample) test or**
- **Mann-Whitney U test**

76

Numerical data: more than two groups

Study: Comparison of >2 sample groups

- **Parametric data**
- Normal distribution
- Means and SD
- Between group variation
- **Use one way analysis of variance (ANOVA)**

- **Non-parametric data**
- Skew or small sample size
- Medians
- **Use Kruskal-Wallis test**

77

Categorical data: a single proportion

Study:

Single sample, each individual possesses characteristic or not (e.g. the number of trainees who are left-handed).

Data is expressed as a proportion. Compare sample proportion to known value.

- **Approximate Normal distribution**
- Proportion
- Estimated SD
- **Use test of single proportion**

78

Categorical data: two proportions

Study: Two groups. Group 1 has characteristic or not. Group 2 has characteristic or not.

- **Groups are independent**
- Data: frequencies
- 2x2 Contingency table
- **Use Chi squared test**
- If frequencies are small use **Fishers exact test**

- **Groups are dependent**
- e.g. same group measured at diff time intervals
- **Use McNemar's test**

79

2X2 Table

80

Categorical data: more than two proportion groups

Study: >Two groups. Group 1 has characteristic or not. Group 2 has characteristic or not, Group 3…. etc

- **Groups are independent**
- Data: frequencies
- 2xC Contingency table

- Use **Chi squared test**
- If frequencies are small use **Fishers exact test**

- If **Groups are ordered** and you want to find a trend

- Use **Chi squared test** for trend

81

PREDICTIVE MODELLING

1. Correlation
2. Linear regression
3. Relative Risk
4. Odds ratios
5. Meta-analysis
6. Survival Analysis

A type of statistical/inferential analysis that not only allows us to describe and compare our sample, but also to make predictions about other samples within the population.

82

1. Correlation

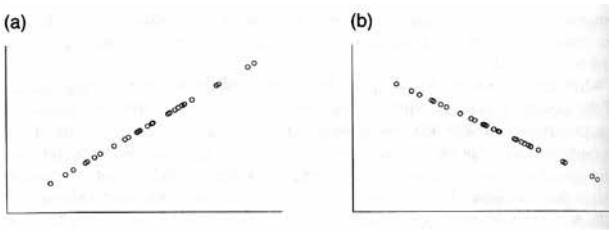
- Assesses the relationship between two variables
  - Example: height and weight, are they related???
- Strength of the association is described by a correlation coefficient: *r (ranges from -1 to 1)*
  - *r* = 0 - .2 low, probably meaningless
  - *r* = .4 - .6 moderate correlation
  - *r* = .8 - 1 very high correlation
- Can be positive or negative
- Pearson's correlation coefficient - parametric
- Spearman's rank correlation coefficient - non parametric

*****Tells nothing about causation!**

Correlation ≠ Causation!

83

Correlation

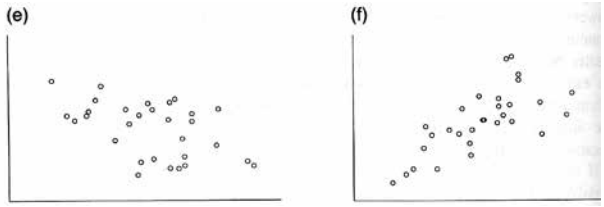


- Correlation is often depicted nicely on graphs like this. For each subject, two variables are plotted against each other, one on the x axis and one on the y axis
- Above graphs show a perfect positive correlation and a perfect negative correlation.

Source: Altman. Practical Statistics for Medical Research

84

Correlation



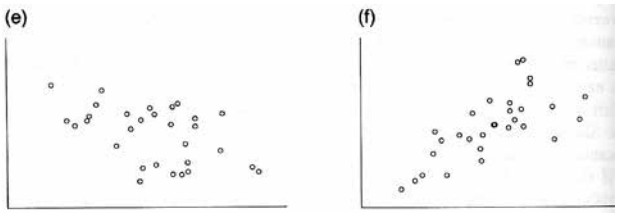
Correlation Coefficient 0

Correlation Coefficient .3

Source: Altman. Practical Statistics for Medical Research

85

Correlation



Correlation Coefficient -.5

Correlation Coefficient .7

Source: Altman. Practical Statistics for Medical Research

86

2. Linear Regression

- Goes a step further than correlation
- Based on fitting a line to data
  - Provides a regression coefficient, which is the slope of the line
    - Y = a + bx
- Use to predict a dependent variable's value based on the value of an independent variable.
  - *Very helpful- e.g. in analysis of height and weight, for a known height, one can predict weight.
- Much more useful than correlation

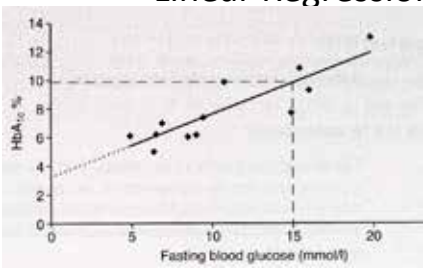
87

2. Linear Regression

- From the concept of correlation, is there a line that approximately fits the distribution of your data?
- More than just saying whether two or more variables are associated, it wants to find out exactly HOW they are related, is there a mathematical formula that relates them.
- This then may allow you to predict further variables from the known variables.

88

Linear Regression



Source: Harris and Taylor. Medical Statistics Made Easy

- Again, we plot our two variables for each subject of our sample. If we look at the points only, we may not see a very good pattern or correlation. But in fact there is a mathematical equation for this line, which is the line that best fits our data.
- If this is our line, we can use the knowledge of one variable to predict the other variable in a given member of the population.

89

Regression

- Types of regression
  - Linear- uses continuous data to predict continuous data outcome
  - Logistic- uses binary data to predict probability of a dichotomous outcome
  - Poisson regression- time between rare events.
  - Cox proportional hazards regression- survival analysis.

90

3. Risk Ratio/Relative Risk

- Typically used in cohort studies
- The risk of having an outcome if you have a risk factor versus the risk of having an outcome if you don't have the risk factor
  - For example: association of smoking with lung cancer

RR = Risk developing disease in exposed group / Risk developing disease in unexposed group

91

Interpreting Risk Ratio

- Risk ratio=1 → no difference in risk
- Risk ratio>1 → increased risk
- Risk ratio<1 → decreased risk
- 95% confidence intervals are usually presented
  - Must not include 1 for the estimate to be statistically significant
    - Example: Risk ratio of 3.1 (95% CI 0.97- 9.41) includes 1, thus would not be statistically significant.

92

4. Odds Ratio

- Case control studies
- Compares the odds compare of having an outcome versus the odds of not having that outcome.
- Different to risk ratio which compares risks of those exposed to those unexposed to specific risk factor.

	CBT	Usual Care (TAU)
Deterioration	3 (13%)	11 (52%)
No Deterioration	20 (83%)	10 (48%)

Rate of deterioration (CBT)	3/23	13%
Odds of deterioration (CBT)	3/20	0.15
Rate of deterioration (TAU)	11/21	52%
Odds of deterioration (TAU)	11/10	1.1

ODDS RATIO of deterioration=0.15/1.1=0.14

The test applied is whether this is different from 1.0

93

Absolute Risk Reduction

	CBT	Usual Care (TAU)
Deterioration	3 (13%)	11 (52%)
No Deterioration	20 (83%)	10 (48%)

Absolute Risk Reduction (ARR) = Deterioration rate (TAU) _ Deterioration rate (CBT)

= 52% – 13% = 39% or 0.39

94

Number Needed To Treat

Absolute Risk Reduction (ARR) = 0.39

Number Needed to Treat (NNT) = 1/ARR = 1/0.39 = 2.56 (~ 3)

- NNT is the number of patients that need to be treated with CBT, compared with treatment as usual, to prevent one patient deteriorating.
- In this case, 3 patients have to be treated to prevent one patient deteriorating.
- NNT can be very useful clinically, because it gives us a real-world sense of the effect of an intervention; e.g. if we were to find that the NNT was 500, and doing CBT was very expensive, from society's point it might not be a useful treatment because we have to treat so many people in order to help only one.

95

5. Systematic Review & Meta-Analysis

- Systematic review performs a review of literature in systematic fashion, determined *a priori*
- Meta-analysis numerically combines data of similar independent studies
  - Estimate an overall/average effect of interest:
  - Uses Relative Risk or Overall Risk
  - Pro: Larger sample and power to detect effects, Efficient, Generalisability
  - Cons: Retrospective data, Missing data, Study heterogeneity, Publication bias

96

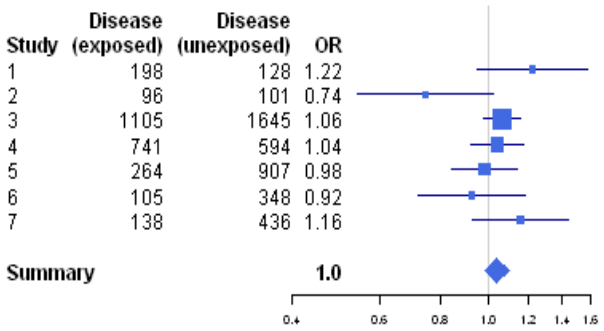
Steps of a Meta-Analysis

1. Define effect of interest.
2. Find appropriate data from published studies (usually proportions e.g. RR or OR).
3. Check statistical homogeneity – use e.g. index I^2
4. Estimate the average effect of interest with CI and perform hypothesis test
5. Null hypothesis – true RR =1
6. Use statistical software to check homogeneity between studies
7. Sensitivity analysis: delete studies and see how it influences plot; particularly if some studies are very different

97

Forrest Plot

to display the results of a meta-analysis



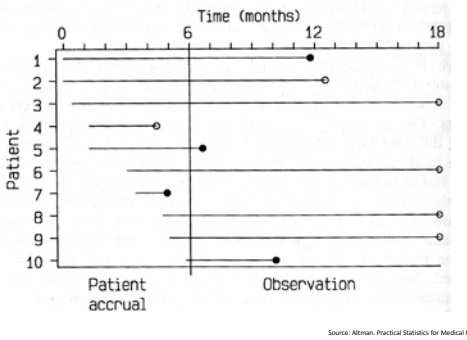
98

6. Survival Analysis

- Evaluation of time to an event (death, recurrence, recover)
- Provides a way of handling censored data
  - Patients who do not reach the event by the end of the study or who are lost to follow-up
- Most common type is Kaplan-Meier analysis
  - Curves presented as stepwise change from baseline
  - There are no fixed intervals of follow-up- survival proportion recalculated after each event.

99

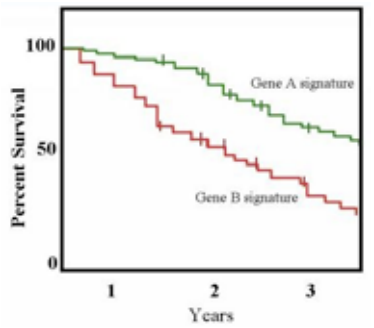
Survival Plot



Source: Altman. Practical Statistics for Medical Research

100

Kaplan-Meier Curve



Source: Wikipedia

101

Statistics Summary

- Understanding basic statistical concepts is central to understanding the medical literature.
- It is NOT important to understand the tests or the underlying maths.
- Need to know when a test should be used and how to interpret its results.

Thank you for your reading and all the best.

102

F2: Practical 6: Statistics

Using online statistics programmes

Find and select a statistics programme that is available for free online use (open access) through an internet search engine (e.g. Google). There are many programmes (e.g. JASP, SOFA, GNU PSPP, Jamovi, IBM SPSS, MacAnova, Invivostat, etc etc) but they all have different strong and weaker points. Sites like GoodFirms (goodfirms.co) try to summarise these (e.g. IBM SPSS is very powerful, MacAnova also works with Linux, Invivostat identifies and removes inaccurate data). Try different programmes with your own experimental data and see which gives you the best understanding of the data and the best visual representation to share your findings (do not try to find a programme to compensate for a poorly conducted study).

Take either data from your own provisional results, or from a pilot study, or take the data from a robust paper you have read, and enter the data. Best is if 2-3 people use the same data in more than one online programme and compare outcomes, then make sure you understand the way the programme does the statistics (you only need to understand the principles of how the quantitative data becomes readable statistics), then discuss which programme would be best suited to your own project.

F.4: Lecture

Qualitative Research Methods & Analysis

J S Dreyer

1

Qualitative Research

Definition: What is Qualitative Research?

= methods to understand human experience of illness, health and treatments.

=How do patients or health care workers find meaning in illness and health, and how it affects their work, life, relationships etc.

2

Qualitative Research Strategies:

1. Ethnography [?] immersing the researcher into a culture to observe a group (e.g. surgical firm).
2. Grounded theory [?] start with no hypothesis, but build the theory from people's stories.
3. Phenomenology [?] study the lived experience of people i.e. How does this disease or treatment affect this person?

3

Qualitative Methods:

1. Interviews
2. Group opinions
3. Questionnaires
4. Case Studies

Data analysis:

1. Test the Validity and Reliability of the data, e.g. through Triangulation.
2. Cronbach alpha is a quantitative method to compare qualitative data.

4

Example

The following slides go through the steps followed to do a qualitative research project on understanding surgical professionalism. The project contributed to achieving a masters degree in surgical education.

5

Qualitative research steps: *Developing a Taxonomy for Surgical Professionalism*

1. Interviews of Surgeon-Educators.
2. Group work (Nominal Group Technique) with surgeons, trainees, nurses and patients.
3. Questionnaires.

Final aim was to try to develop a quantitative measuring tool to assess professionalism in surgeons.

6

Building Taxonomy from Literature review:

How to classify Professionalism Characteristics

Arnold & Stern (2006)

Figure 2-1 A definition of professionalism.

7

Interviews:

When typed it looks like this

- FD: I am here with Mr. Ian Ritchie in Stirling
- Ian we have discussed some of these aspects of professionalism. I just want to ask you firstly, in the light of what we have discussed and what you have seen, what do you understand under the concept of professionalism in surgery in basic terms?
- IR: It is very wide and it is difficult to define it accurately, but I guess professionalism means applying scientific knowledge for the benefit of the patient, but also taking into account what the patient's needs and circumstances are. It also involves behaviours in relation to the patient that take account of the patient's vulnerabilities and their, because of their lack of knowledge, perhaps their inability to influence the outcome. So, it involves respecting the trust that has been given to you. I think that is probably most of it.

8

Next: Colour coding of essential words based on categories from literature review:

- Excellence = red
- Accountability = green
- Humanism = brown
- Altruism = dark blue
- Ethics = orange
- Relationships = light blue

9

Colour coding transcribed text

It is very wide and it is difficult to define it accurately, but I guess professionalism means applying scientific knowledge for the benefit of the patient, but also taking into account what the patient's needs and circumstances are. It also involves behaviours in relation to the patient that take account of the patient's vulnerabilities and their, because of their lack of knowledge, perhaps their inability to influence the outcome. So, it involves respecting the trust that has been given to you. I think that is probably most of it.

10

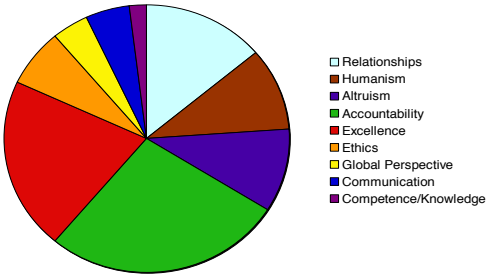
Results:
After colour coding all Interview transcripts

Characteristics rated highest were:

- Excellence (37): high clinical standards; judgement.
- Accountability (48): honesty; trustworthiness; transparency; insight into own strengths & weaknesses
- Humanism (17): respect, compassion, fairness
- Altruism (18): put patients' needs first
- Ethical behaviour (12): high moral standards; probity.
- Managing inter-personal relationships (26): humility; avoid negativity; be realistic in expectations of trainees; accept others doing things differently; be prepared to learn from a variety of people.

11

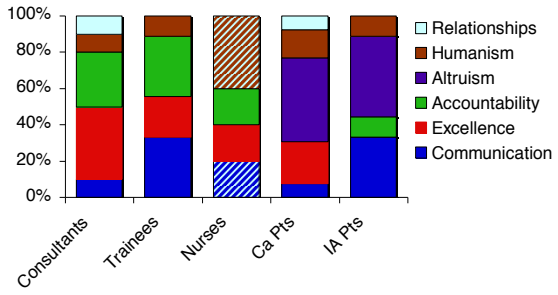
Interviews: Distribution of Characteristics (n=178)
Graphs give better visual display



12

Nominal Group Technique Results:

Distribution of Characteristics (using same colours for the same characteristics as identified in interviews and literature review)



13

What happened next

- We had 112 characteristics in new Taxonomy:
  - 50 from Literature
  - 62 from Interview + NGTs
- We then decided to refine these down to the most valued 20 characteristics of professionalism through a series of questionnaires that refined the data through each cycle.
- This is known as the "**Delphi process**"

14

Questionnaire1

- To surgeons and trainees in Scotland AND to other doctors, nurses and patients in Dumfries & Galloway
- Asked to select 3-7 characteristics in each category.
- 704 replies (M:F =1:2)
  - Doctors 163 (24%)
  - Nurses 274 (40%)
  - Patients 247 (36%)
- Selected 58 characteristics

15

Questionnaire 2:

- To consultant and trainee surgeons in Scotland.
- Used the 58 characteristics from Questionnaire 1, written as questions.
- Asked to mark "the professionalism characteristics you would most value in a surgical trainee (maximum 20)"

16

Final outcome

Integrating all results has produced a combined list of 20 characteristics, e.g:

- Excellence

"Can the trainee deal effectively with complexity and uncertainty?"

- Accountability

"Is the trainee honest when errors occur?"

17

- Humanism

"Is the trainee open and honest when taking consent?"

- Altruism

"Does the trainee value and maintain confidentiality?"

- Interpersonal Relationships

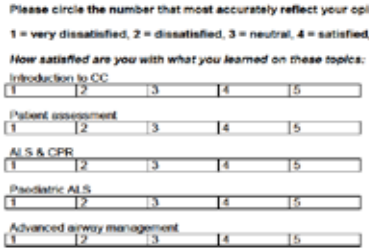
"Is the trainee prepared to learn from a variety of people?"

18

Another example of qualitative research: Collecting feedback from a course or other teaching event

19

Feedback/Course Evaluation Questionnaires



Please circle the number that most accurately reflect your opinion where

1 = very dissatisfied, 2 = dissatisfied, 3 = neutral, 4 = satisfied, 5 = very satisfied.

How satisfied are you with what you learned on these topics:

Introduction to CC

1	2	3	4	5

Patient assessment

1	2	3	4	5

ALS & CPR

1	2	3	4	5

Paediatric ALS

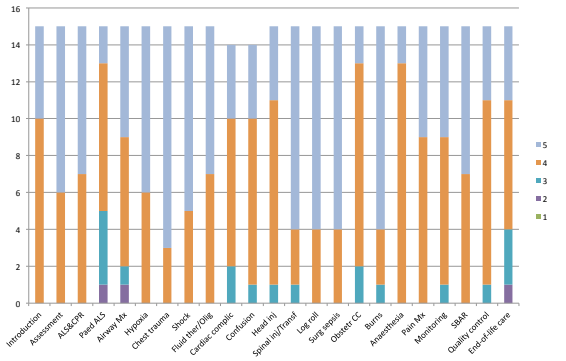
1	2	3	4	5

Advanced airway management

1	2	3	4	5

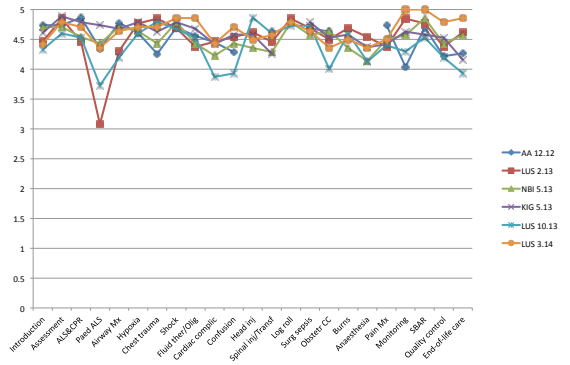
20

Feedback scores compiled vertically



21

Feedback scores compiled horizontally



22

Quantifying feedback comments

What was very good?
Attitude/participation of faculty (x7)
Introductions to make everyone comfortable.
Tutorials/Small group teaching (x6)
Organisation
Group interactions
Practical sessions/demonstrations (x5)
Concepts simplified (x2)
Time keeping (x3)
SBAR
Quality control/patient safety
What could be better?
Proper handbook/More notes/make ptolemy articles more accessible/ available earlier (x6)
More practical demonstrations
Use videos/diagrams more (x2)
Stronger/safer chairs (x2)
Time allocation for written test too short/give test on paper (x2)
In monitoring, discuss invasive methods as well
More time
Include ENT
Other comments?
Time allocation for sessions too short.
Very relevant material (x2)
Increase the length of the course
Too much attention on assessment impairs learning opportunities

23

Quantifying feedback comments

What was very good?
Attitude/participation of faculty (x7)
Tutorials/Small group teaching (x6)
Practical sessions/demonstrations (x5)
Concepts simplified (x2)
Time keeping (x3)

What could be better?
Proper handbook/More notes/make ptolemy articles more accessible/ available earlier (x6)
Use videos/diagrams more (x2)
Stronger/safer chairs (x2)
Time allocation for written test too short/give test on paper (x2)

Other comments?
Time allocation for sessions too short/Increase the length of the course
Very relevant material (x2)

24

Conclusions

1. Read what others have done.
2. Be innovative
3. Make it simple.
4. Read extra

25

F.5: Lecture

Other dimensions in patient data analysis

J S Dreyer

1

Traditional outcome measures

- Mortality
  - Real vs expected deaths
  - Is it always relevant?
  - Does it mean “cure” or “survival”?
- Morbidity
  - Immediate/short term e.g. post-op complications
  - Medium term e.g. cancer recurrence
  - Late e.g. Disability (DALYs)

2

Other outcome measures to consider:

- Health related Quality of Life (HRQOL)
- Patient satisfaction
- Quality of care/Patient safety

3

Quality of Life indicators

- WHO define health as not just the absence of disease but as a state of complete physical, mental and social well-being.
- QoL indicators try to capture patients’ own perceptions of their health and ability to function in their daily lives.
- Assessment instruments should therefore be self-administered.

4

Example of survey tool [1]

- General Health: SF-36
  - Physical function
  - Role limitations from physical problems
  - Bodily pain
  - Energy/fatigue
  - Mental health
  - Role limitations from emotional problems
  - General health perceptions

5

Example [2]

- "FACT"

= Functional Assessment of Cancer Therapy

1. Physical
2. Social/Family
3. Relationship with doctor
4. Emotional
5. Functional

6

Recommendations on including HRQOL outcomes in clinical research

1. Use validated HRQOL instruments
2. Use correct instrument for your study focus
3. Get multidisciplinary input on QoL questions early (in study design)
4. Assess QoL longitudinally (over time, not as single snapshots)
5. Use controls

7

Patient Satisfaction Outcomes

- Patient satisfaction often depends more on the "care" than the "cure"
- e.g. in cancer treatment:
  - Cancer control
  - HRQOL
  - Side effects of treatment e.g. erectile dysfunction
  - Recovery quality and time
  - Ability to perform daily activities
  - Psychological wellbeing
  - Financial outcomes
  - Quality of death

8

Quality of Care/Patient Safety in Surgical Research

- High quality care "consistently contributes to better quality of life and/or a longer life" (AMA)
- Quality of care measured in:
  1. Structure
  2. Process
  3. Outcomes

9

1. Structure

- Administrative structure e.g.
  - Number of beds
  - Nurse-to-patient ratio

  (Lay press like to report on this)
- Clinical structure e.g.
  - Hospital volume
  - Surgeon procedure volume

  (BUT volume of surgery is not a surrogate for quality of care)

10

2. Process

- Much more difficult to measure how processes affect patient outcomes:
- Successful examples = use of
  - WHO Safe Surgery Checklist
  - Pre-operative antibiotics
  - VTE prophylaxis
- Processes of Care are very good for AUDIT (finding out "what happens?") leading on to RESEARCH (looking at "how?" or "why?")

11

Example: Audit in Safe Surgery:

(a) Audit the process of using a Checklist (SSC):

e.g: -how the SSC is implemented and/or accepted
-practical problems in using different parts of the SSC
-developing a protocol for swab and sharps counts.

(b) The effects of the SSC [?] Specific Outcomes

e.g: -wound infection or DVT incidence
-drug allergic reaction incidence
-delay in getting blood to theatre

(c) Qualitative research on how the SSC affects practice:

e.g: -how did introducing the SSC affect theatre teamwork or communication.
-how difficult was it to change theatre practice.

12

3. Outcomes

= effect of surgical care on the health status of patients and populations:

- Patients:
  - Clinical
  - Physiological
  - Psychological
  - Social
- Populations: e.g.
  - Population screening for bowel cancer
  - WHA resolution 68.15 to include emergency and essential surgery as part of global universal health care.

13

Summary

1. Think of patients' Quality of Life when assessing treatment benefit.
2. Life has 100% mortality; it is the quality of your years, not the quantity that counts most for most.
3. Use established QoL tools; the established ones are available through the internet (remember to give recognition to where you downloaded from).
4. Think and read widely.
5. Put yourself in patients' shoes.

14

F7: Practical 7

How to analyse data from your own study

Discuss potential analysis methods for study proposals in your group e.g. what type of data (quantitative, qualitative) do we expect to get, therefore what tests to use; can we find a statistician to help; which internet available statistics programme can we use? do we need to quantify any qualitative data? Let the principal investigator for each study present their ideas for 5-10 minutes, and then have a 10-15 minute discussion; end by writing down 3 key points on data analysis for each study proposal.

G1: Lecture

Presenting your results

Frieda Elsje Dreyer
Jonathan AF Hannay

1

Learning outcomes

- Preparation of results
- Presenting your findings
  - Poster
  - Platform
- Summary

2

Preparation of results

- Do you understand your own results?
- Does it answer the original question (primary outcome)?
- Does it answer any other questions (secondary outcome)?

- Are they
1. Clean
2. Correct
3. Clear
4. Concise
5. Comprehensible
6. Communicable

3

Presenting your findings

Structure your findings and results relevant to your audience and venue

- **Local departmental meetings**
  - Does it identify the need for audit, change in local guidelines and practice, or highlight need for a Quality Improvement project?
- **Scientific meetings**
  - Can be regional, national or international
  - May be specialty forums
  - Does it highlight an interesting topic, new research or identify the need for standardising practice nationally/ updating national guidance?
- **Journal**
  - Generic or specialty focused
  - In the form of a letter/research article / systematic review
- **Government body/ Department of Health**
- **Club or charity**

4

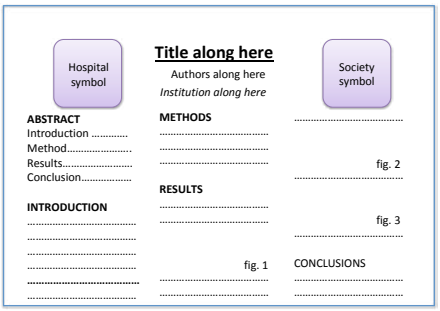
Poster presentation

Practical tips:

- Read the instructions from the organiser
- Including size of poster/ number of images, charts, words allowed/ landscape or portrait
- Type of poster i.e. electronic (iPoster), interactive, paper
- Do you need to print it yourself or does the organiser offer a printing facility usually at a increased fee
- Can you print on canvas to allow easier transport?
- Does your hospital or trust have a generic template you can use, these may appear more professional especially if a number of scientists from the same trust present at the forum.
- Add the professional body/society's and your hospital's symbol to the poster as illustrated
- Create your poster on PowerPoint
- Before printing, save a PDF file and check your margins/ charts fit on the page to avoid words being cut off mid sentence
- Consider a plain background and font easy to read with appropriate size and spacing
- Grab attention with a catchy title, charts and images
- Acknowledge all your contributors even if they don't make the shortlist for named authors

5

Poster presentation



6

Platform presentation

- A platform presentation is a 10-15 minute oral presentation of an original research project or paper followed by 5 minutes of question time, moderated by a chair.
- It provides an opportunity to present your ideas to a large audience and to highlight key elements of your research that are unique, novel or contribute new knowledge to the field.
- Check instructions for speakers including time limit, venue, type of audience.
- The type of audience will allow you to tailor your focus of the presentation
- Begin your presentation with a statement of any conflicts of interest or disclosures
- Clearly state your aim
- Your presentation should follow a logical order, the format of a traditional scientific abstract and include the following sections:
  - Background
  - Objectives
  - Aims
  - Methods
  - Results
  - Limitations
  - Conclusions

7

Platform presentation

Tips for speakers:

- Check your slides for format, spelling, grammar and have the most up to date version ready
- Speak clearly
- Keep to time
- Make eye contact with your audience
- Know your subject
- Provide a summary slide
- Acknowledge appropriate people
- Invite questions with the chairman's permission
- Listen to the question, repeat it back if necessary and pause before answering

8

Summary

1. Research can be presented in various formats
2. Know your subject
3. Know your audience
4. Be clear
5. Diagrams and charts are good visual aids
6. Keep animations to a minimum as these can distract from you message
7. Prepare and practice
8. Listen to the question

9

Summary

- Practice, practice, practice….
- Start in your own department, then your own hospital or university, then a regional or national conference before going to an international conference. That way you will build your confidence and refine your material and presentation style.
- When you have a bad day, review and go again (it is like playing sport).

10

G2: Lecture

How to write a paper

Stephan B Dreyer

1

Contents

- What is a research paper?
- How to structure your paper
- What language is appropriate
- Formatting

2

What is a research paper

- Structured and evidenced story
- Clear problem or question
- Possible solution
  - (or attempt at a solution)
- Provides a springboard for further research and clinical development

3

Example

SPECIAL ARTICLE

A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population

4

How to structure your paper

- Introduction
  - Aims
- Methods
- Results
- Discussion
  - Conclusion

5

Introduction

- Introduce the clinical problem and why it is important to patients (and or science)
- Lead the reader to believe that this is an important problem to address
- Get their attention early – 1st and 2nd sentence *(see Example: 1st arrow in paper slide)*

6

Introduction

- Background on what's known on the topic
  - Concise
  - Relevant to the research question
- Evidenced by most recent and relevant literature
  - Not an extensive literature review
- Lead the reader to your question
- Aims
  - Be clear and defined

(See examples indicated by arrows in paper slide [next])

7

8

Materials and Methods

- What you did
- How you did it
  - Outcome measures
- Why you did it this way

AND

- Statistical analysis

9

10

11

12

Results

- Describe the study cohort
  - Table
- Lay out each question
- Do not repeat contents in tables or figures in text
- Clear legends and labels for tables and figures
  - Make it easy for the reader to understand

13

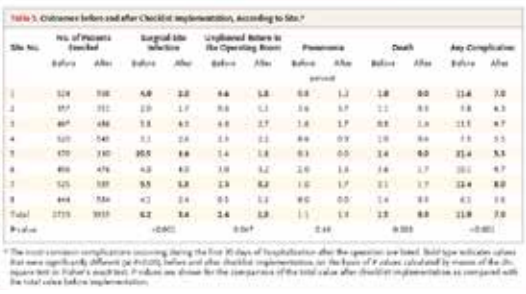
Results

RESULTS

We enrolled 3733 patients during the baseline period and 3955 patients after implementation of the checklist. Table 4 lists characteristics of the patients and their distribution among the sites; there were no significant differences between the patients in the two phases of the study.

The rate of any complication at all sites dropped from 11.0% at baseline to 7.0% after introduction of the checklist ($P<0.001$); the total in-hospital rate of death dropped from 1.5% to 0.8% ($P=0.003$) (Table 5). The overall rates of surgical-site infection and unplanned reoperation also declined significantly ($P<0.001$ and $P=0.047$, respectively). Operative data were collected by the local data collector through direct observation for 37.5% of patients and by unobserved clinical forms for the remainder. Neither the presence nor

14

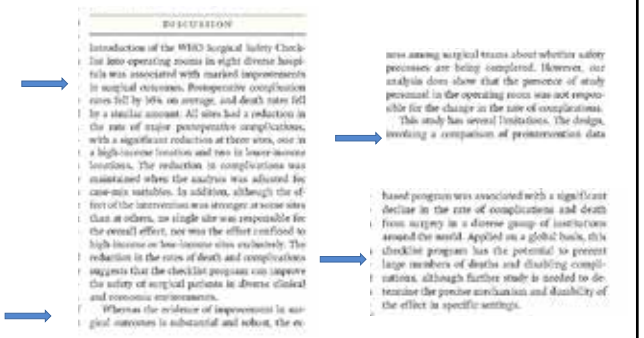


15

Discussion

- 1st Paragraph: What you found
- 2nd Paragraph: What other studies found
- 3rd Paragraph: What does your study add
- 2nd last: Weaknesses
- Last: Conclusions and next steps

16



DISCUSSION

Introduction of the WHO Surgical Safety Checklist into operating rooms in eight diverse hospitals was associated with marked improvements in surgical outcomes. Postoperative complication rates fell by 36% on average, and death rates fell by a similar amount. All sites had a reduction in the rate of major postoperative complications, with a significant reduction at three sites, one in a high-income location and two in lower-income locations. The reduction in complications was maintained when the analysis was adjusted for case-mix variables. In addition, although the effect of the intervention was stronger at some sites than at others, no single site was responsible for the overall effect, nor was the effect confined to high-income or low-income sites exclusively. The reduction in the rates of death and complications suggests that the checklist program can improve the safety of surgical patients in diverse clinical and economic environments.

Whereas the evidence of improvement in surgical outcomes is substantial and robust, the evi-

ness among surgical teams about whether safety processes are being completed. However, our analysis does show that the presence of study personnel in the operating room was not responsible for the change in the rate of complications.

This study has several limitations. The design, involving a comparison of preintervention data

based program was associated with a significant decline in the rate of complications and death from surgery in a diverse group of institutions around the world. Applied on a global basis, this checklist program has the potential to prevent large numbers of deaths and disabling complications, although further study is needed to determine the precise mechanism and durability of the effect in specific settings.

17

Abstract

- Write this last
- Short summary of the most important findings
- Stick to the key findings only
- Be concise
- Stick to word limit

18

ABSTRACT

19

Language

- Clear and concise
- Use short sentences
  - Matter of fact
- Use appropriate language for the audience
  - Clinicians vs Scientists

20

How to get it published?

- Pick the right journal (have a 1st, 2nd, 3rd choice for publication)
- Format it to the journal's specs
  - Read *(lots of)* papers from that journal
- Use reference managers (e.g. ENDNOTE)
- Revise and revise and revise
- Keep trying

21

Just DO it

And Good Luck…
Nothing worthwhile comes easily.

22

G.4: Practical 8 (presenting and writing)

How to prepare for presenting and/or writing up your own project

You can practice presentation and writing before you have completed your own research project as long as you have a small group of research enthusiasts that support each other. Things you can do include:

a. Prepare an oral presentation on a paper you have read as if it were your own study and you have to share the results with your peers.
b. Prepare an electronic poster on a paper and present at your meeting (you can do the poster in PowerPoint and project from your laptop, so no expenses necessary).
c. When you have read 5-10 good for your study proposal, summarise these into an oral presentation to your study group or your own hospital department. You can summarise these under the IMRD headings you will use to write your own paper.
d. Write a summary abstract for yourself on these background papers, as if you are going to submit this as an abstract for a conference.
e. Use (c) and (d) as the core information to write up the literature review for when you want to publish your own study.
f. Write up a summary of the methods you want to use for your own study and present to your peer study group. Welcome any positive criticism to improve your study.

H1: Lecture

Overcoming Barriers in Research

Ainhoa Costas-Chavarri

1

Potential barriers

1. How to do research with limited resources.
2. How to form/join research collaborations.
3. How to make research work in my own institution.

2

So you want to do research

- Where to start?
  - Design your research question
  - Use what you have learned in this course
- How do I keep going?
  - Confronting challenges
  - Identifying resources
  - Developing a plan

3

Challenges

Getting research topics	80%
Lack of protected research time	70 %
Financial burden	64%
Inadequate research training	54%
Low value attached to research output	50 %
Inadequate mentorship and supervision	50%

ELOBU Emmanuel Alex,
MD(Dar), MMed Surgery (MUK)
FCS-ECSA, Dip ACREM
13th COSECSA AGM Addis Ababa, Ethiopia

4

Challenges

- Ideas

- Time
- Funding
- Experience
- Human/People
- Infrastructure/Facilities

5

IDENTIFYING RESOURCES

6

WHAT Resources DO YOU NEED FOR YOUR STUDY?

- 1. People
- 2. Facilities
- 3. Funding
- 4. Time

7

People

- Identify mentors
  - Faculty advisor or supervisor
  - Can be local or abroad
- Where to find mentors
  - Local or international leaders
  - Conferences
    - Giving presentations
    - Experts on panels
  - Publications
    - Authors of relevant articles

- Make connections!

8

PEOPLE (2)

- Identify team or assistants
  - Students
  - Postgraduates
  - Colleagues in other fields
    - Public Health
    - Statistician
- Think "Collaborations"

9

What is a Research Collaborative?

- A network of people interested in and collaborating in research studies
  - Regional
  - National
  - International
  - Across specialties
- AIM =

  To "Deliver high quality, multi-centre clinical research that will change clinical practice"

10

Why Research collaboratives?

- Establish contacts
- Combine research experience
- Improve research
  - 'Rare' diseases or procedures
  - Larger volume of patients
  - Explore Differences
    - E.g. **'Practice Variability'** that leads to **'differences in outcome'**
  - Multi-centre
  - Multi-national
- Changes in practice

11

Examples

http://nationalresearch.org.uk

12

How to get involved

- GlobalSurg
  - GlobalSurg2: determining the epidemiology of surgical site infections after abdominal surgery
  - >12,000 patients
  - 343 hospitals
  - 66 countries
  - Multiple other stu
  - Students and residents encouraged to participate
    - Gain experience and make connections
- Incision
  - International Student Surgical Network
- GSSA
  - Global Surgery Student Alliance
- WiSA
  - Women In Surgery Africa
  - Let's collaborate initiative
- COSECSA
- Start your own!

13

Infrastructure/facilities

Where will your study take place?

- Clinical Sites?
- Simulation Center?
- Office/Workspace?
  - Storage?

What do you need for your research?

- Equipment
- Medications
- Materials

14

Funding

- Create a budget
  - Allowance for assistants
  - Cost of infrastructure/facilities
  - Equipment/materials/medications
    - Laptop or Software (Statistics!)
  - Cost of IRB/Ethics approval
  - Transport/Travel
  - Communications/Airtime/Data

15

Funding

- Global interest in Surgical / NCD research
  - Take advantage
- Look for funding awards
  - COSECSA / RCSI
  - RCSEd
  - ASCO

16

Funding

- Most clinical projects require no funding
- Be flexible and creative!

17

DEVELOPING A PLAN

18

Time – What do you need to do?

- Writing the proposal
  - Include your budget
- IRB/Ethics approval
  - More complex if research is in Multiple sites
- Data collection
- Data synthesis and analysis

19

Creating a timeline

- Create an action plan
- Create a schedule for this plan
- Start with the end and work backwards:
  - Thesis due date
  - Conference deadline
  - Grant deadline

20

Creating a timeline 

- Convert your timeline into a checklist
- Set reminders for yourself
- Check your timeline often and ask for research time when you need it
  - Divide and share work among team members

21

Anticipate barriers and delays!

- IRB/Ethics approval delays
- Problems with patient recruitment
- Data collection problems
- Time it actually takes to complete each step

22

Get started!

- Determine resources you will need
- Identify mentors, team, collaborators
- Define your infrastructure and material needs
- Create a budget
- Develop a timeline/plan

23

You GO for it!

24

H.2: Practical 9

The value of a local research support group

Try to find likeminded colleagues who are your peers, e.g. other surgical residents that you work with in the same hospital, or with residents from other departments, or other health care practitioners that are interested in clinical research (laboratory staff, nutritionists, public health doctors [who often know more about statistics and epidemiology than surgeons] etc). The trick is to think widely. Invite 1-3 supportive consultants to attend, even if they might be very senior academics. Most professors love young researcher to come up with new ideas and would like to know about these early. Find a time and place to meet regularly when it would suit everybody to meet most of the time. Do not overdo the meetings; monthly meetings will be more sustainable than starting with weekly meetings because you are enthusiastic. Keep very short minutes of you meetings so you can remind people what you discussed previously and have an agenda for every meeting, so people can prepare to discuss their specific project successes or problems. Invite speakers to come and talk about difficult issues e.g. a statistician from the local university (irrespective whether medical or not), somebody who has achieved research success in a different field, someone who can advise on publication (e.g. an editorial board member of a medical journal). Most important is to be open, supportive, non-threatening and non-defensive. Criticism must always be positive and aimed at the contents the group discusses; never be critical of researchers in person, even if they do not attend the meeting. Make every meeting a learning experience and end the meeting with a short list of 3-5 learning points from that specific meeting. Then share a meal or a few drinks or go home if you have a family and don't talk about research when you have left the meeting.

Appendix 1:

Appendix 1 contains a number of journal papers for critical reading and analysis. These papers are all available for free download through e.g. Google Scholar and copies are available in the course handbook. It works best if 1-2 persons read and summarise one paper and then present their findings to colleagues within a study group. If you give 10 minutes for each presentation and 5 minutes for discussion you can go through 5 papers in under 90 minutes. That should give you sufficient confidence to be a critical future reader.

OPEN

Nonoperative Treatment With Antibiotics Versus Surgery for Acute Nonperforated Appendicitis in Children

A Pilot Randomized Controlled Trial

Jan F. Svensson, MD,† Barbora Patkova, MD,*† Markus Almström, MD,*† Hussein Naji, MD,*† Nigel J. Hall, MD, PhD,‡§ Simon Eaton, PhD,‡‖ Agostino Pierro, MD, PhD,§ and Tomas Wester, MD, PhD*†*

Objective: The aim of this study was to evaluate the feasibility and safety of nonoperative treatment of acute nonperforated appendicitis with antibiotics in children.

Methods: A pilot randomized controlled trial was performed comparing nonoperative treatment with antibiotics versus surgery for acute appendicitis in children. Patients with imaging-confirmed acute nonperforated appendicitis who would normally have had emergency appendectomy were randomized either to treatment with antibiotics or to surgery. Follow-up was for 1 year.

Results: Fifty patients were enrolled; 26 were randomized to surgery and 24 to nonoperative treatment with antibiotics. All children in the surgery group had histopathologically confirmed acute appendicitis, and there were no significant complications in this group. Two of 24 patients in the nonoperative treatment group had appendectomy within the time of primary antibiotic treatment and 1 patient after 9 months for recurrent acute appendicitis. Another 6 patients have had an appendectomy due to recurrent abdominal pain (n = 5) or parental wish (n = 1) during the follow-up period; none of these 6 patients had evidence of appendicitis on histopathological examination.

Conclusions: Twenty-two of 24 patients (92%) treated with antibiotics had initial resolution of symptoms. Of these 22, only 1 patient (5%) had recurrence of acute appendicitis during follow-up. Overall, 62% of patients have not had an appendectomy during the follow-up period. This pilot trial suggests that nonoperative treatment of acute appendicitis in children is feasible and safe and that further investigation of nonoperative treatment is warranted.

Keywords: antibiotics, appendicitis, children, randomized controlled trial, surgery

(*Ann Surg* 2015;261:67–71)

Acute appendicitis is the most common disease requiring emergency surgical treatment in children. Traditionally, the standard treatment of acute appendicitis has been appendectomy. However, there is growing interest in nonoperative treatment of acute nonperforated appendicitis with antibiotics. Several randomized controlled trials (RCTs) have been performed in adults and these have also been subjected to meta-analysis. Data suggest that antibiotic treatment may be an effective treatment modality for adults with acute nonperforated appendicitis and that approximately 75% of patients may not need appendectomy at all, either during initial illness or during the first year of follow-up.[1] However, a recent Cochrane review concluded that further well-designed RCTs were needed.[2]

In children, although there have been several studies of initial conservative treatment of *perforated* appendicitis,[1,3,4] data on conservative treatment of *nonperforated* acute appendicitis in children are scanty. The only comparative published study was retrospective and had unclear diagnostic and treatment criteria.[5] Of note, there have been no RCTs investigating nonoperative treatment of acute nonperforated appendicitis in children.

As a prelude to a large RCT investigating the efficacy of nonoperative treatment of acute nonperforated appendicitis in children, we designed a pilot RCT to inform our future planned study. The objectives of this pilot study were to (1) evaluate the feasibility of recruiting children with acute appendicitis to an RCT comparing nonoperative treatment with appendectomy, (2) evaluate the safety of nonoperative treatment with antibiotics of acute nonperforated appendicitis in children, and (3) generate pilot data to inform our future planned efficacy study.

METHODS

Trial Design

This was a pilot trial comparing nonoperative treatment (antibiotics) and surgery for acute nonperforated appendicitis in children. The diagnosis was made with the combination of clinical findings and imaging. All children underwent abdominal ultrasound scan, and a computed tomographic (CT) scan was performed when there was diagnostic uncertainty. Age, sex, duration of symptoms, body temperature, and C-reactive protein, white blood cell, and neutrophil concentrations at admission were recorded.

Participants

All children between 5 and 15 years of age with a clinical diagnosis of acute appendicitis that before the trial would have been subjected to an appendectomy, including those with an appendicolith, were eligible. Exclusion criteria were (1) suspicion of perforated appendicitis on the basis of generalized peritonitis; (2) an appendiceal mass, diagnosed by clinical examination and/or imaging; or (3) previous nonoperative treatment of acute appendicitis.

Study Setting

The study was conducted at the Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden. This

From the *Department of Pediatric Surgery, Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden; †Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden; ‡Surgery Unit, UCL Institute of Child Health, London, United Kingdom; §Division of General and Thoracic Surgery, The Hospital for Sick Children, Toronto, Ontario, Canada; and ‖Department of Pediatric Surgery, Great Ormond Street Hospital, London, United Kingdom.

The trial is registered at ClinicalTrials.gov: No. NCT01572558.

Disclosure: Supported by the Crown Princess Lovisa's Foundation and the Hirsch Foundation grants. The funders had no role in study design, data collection/analysis, and decision to publish the manuscript.

The authors declare no conflicts of interest.

Reprints: Jan F. Svensson, MD, Department of Pediatric Surgery, Astrid Lindgren Children's Hospital, Karolinska University Hospital, Q3:03, 17176 Stockholm, Sweden. E-mail: jan.f.svensson@karolinska.se.

ISSN: 0003-4932/14/26101-0067

DOI: 10.1097/SLA.0000000000000835

is the only hospital with a pediatric surgical service within the greater Stockholm area and serves a population of approximately 2.5 million inhabitants.

Interventions

Enrollment in this study was after the attending pediatric surgeon had made a diagnosis of acute appendicitis, the patients and their family had received oral and written information regarding the trial, and the patients and their family had provided written informed consent to participate. Children with acute nonperforated appendicitis were randomly allocated to either appendectomy or nonoperative treatment with antibiotics. All patients allocated to surgery received preoperative antibiotic prophylaxis with 20 mg/kg of metronidazole. Further antibiotic treatment in this group depended on the severity of appendicitis in accordance with institutional practice. Cases of simple or phlegmonous appendicitis received no further antibiotics, those with gangrenous appendicitis received 24 hours of intravenous trimethoprim/sulfametoxazol/metronidazole, and those with perforated appendicitis received at least 3 days of intravenous trimethoprim/sulfametoxazol/metronidazole, depending on clinical course. The modality of surgery (open or laparoscopic) was not stipulated in the trial protocol.

Children allocated to antibiotic treatment were given intravenous meropenem (10 mg/kg × 3 per 24 hours) and metronidazole (20 mg/kg × 1 per 24 hours) for at least 48 hours. Once the child was clinically well and tolerating oral intake, the treatment was changed to oral ciprofloxacin (20 mg/kg × 2 per 24 hours) and metronidazole (20 mg/kg × 1 per 24 hours) for another 8 days. The protocol stipulated that children should be kept nil by mouth for the first 24 hours, but in practice, we found this hard to enforce as children were clinically well and often demanded to drink and eat earlier. Criteria for discharge were established a priori and applied to both treatment groups equally. They were as follows: afebrile for 24 hours, with or without oral antibiotics, adequate pain relief on oral analgesia, tolerating a light diet, and mobile.

OUTCOMES

The primary outcome was the proportion of children in each group achieving "resolution of symptoms without significant complications." This outcome was chosen as it was applicable to both treatment arms and also because this constitutes a pragmatic goal for a patient coming to the hospital with appendicitis. Significant complications were defined as length of stay more than 7 days, abscess formation, the need for surgery within 48 hours in the antibiotic group, recurrence of appendicitis within 3 months, and negative appendectomy. Secondary outcomes measured were time from randomization to discharge, complications (wound infection, wound dehiscence, diarrhea, etc), and recurrent appendicitis within 1 year of randomization. To monitor children recruited into the study and to allow collection of a full data set, all participants were seen in the outpatient clinic at 4 to 6 weeks after discharge, with further follow-up visits at 3 and 12 months after randomization. Because we encountered difficulties getting the patients to return to the outpatient clinic at 1 year after randomization, we accepted a telephone interview with one of the parents as 1-year follow-up. The specific purpose of this 1-year follow-up was to identify episodes of recurrent appendicitis and any children who had undergone appendectomy due to recurrent symptoms or parental request at another center. We do not believe that conducting this review by telephone as opposed to in person results in significant bias for this particular outcome. As a result, we had 1-year outcomes on all patients enrolled in the trial.

Total cost of treatment was calculated in a pragmatic way, as reimbursement methods differ between different countries and systems. Total cost per participant was calculated as a fee per day of in-hospital care, a fee for use of the operating room, and the cost of a course of intravenous and oral antibiotics for the nonoperative treatment group. Total costs that include cost for the initial hospital stay for both treatment groups and cost for any additional admission as applicable are presented.

Sample Size

As this was a pilot trial, we did not perform a power calculation. On the basis of our yearly caseload of approximately 400 cases and estimated recruitment of one third of eligible cases, we aimed to enroll 50 patients within a 6-month period.

Randomization

Allocation to groups (1:1 ratio) was made via weighted minimization at the time of enrollment in the study using the following criteria: age (5–10 years or 11–15 years), sex (male or female), and duration of symptoms (<48 or ≥ 48 hours). All factors were weighted equally. Randomization was performed using a computer-based randomization program (Simin v 6.0; Institute of Child Health, London), which allowed complete concealment of randomization sequence.

Blinding

As this was a pilot trial comparing surgery and nonoperative treatment with antibiotics, it was not considered possible or ethical to blind patients, parents, or surgeons.

Statistical Methods

Data are presented as the proportion of participants or median (range). Data were compared using the Mann-Whitney *U* test or the Fisher exact test as appropriate, using IBM SPSS Statistics version 22. This trial is reported in accordance with the CONSORT statement.[6,7]

Ethical Approval

The study was approved by the Regional Ethics Review Board (reference No. 2011/1234-31/4).

RESULTS

The trial opened on February 7, 2012, and the final participant was enrolled on October 17 the same year. One-year follow up for the cohort was completed on October 25, 2013. During the trial period, 225 children with a clinical diagnosis of acute appendicitis that before the trial would have been subjected to an appendectomy were seen at our institution. A total of 174 children were not enrolled in the trial for reasons shown in Figure 1. In addition to the defined exclusion criteria and parental nonagreement to participate, 2 children were excluded on the basis of CT findings, one with a suspicion of a carcinoid tumor of the appendix and one in whom it was impossible to differentiate between appendicitis and a Meckel diverticulitis. Overall, 52 of the 129 children (40%) whose parents were asked whether they would consent to their child being in the trial agreed. After agreement to participate, there was failure of the computer randomization program affecting 1 case (this child was not included in the study) and in 1 case parents withdrew consent to participate in the study after allocation of treatment. This child was withdrawn from the study. To account for these 2 cases, additional participants were recruited to reach the target sample size of 50.

Participants had similar demographic and admission characteristics both to those children whose parents declined participation and to those children who were not invited to participate (Table 1) except that the proportion of children with symptom duration of less than 48 hours was significantly lower in the group of participants who were not offered to participate in the trial. The reason for this is unclear, although it is possible that surgeons felt that there was a clearer

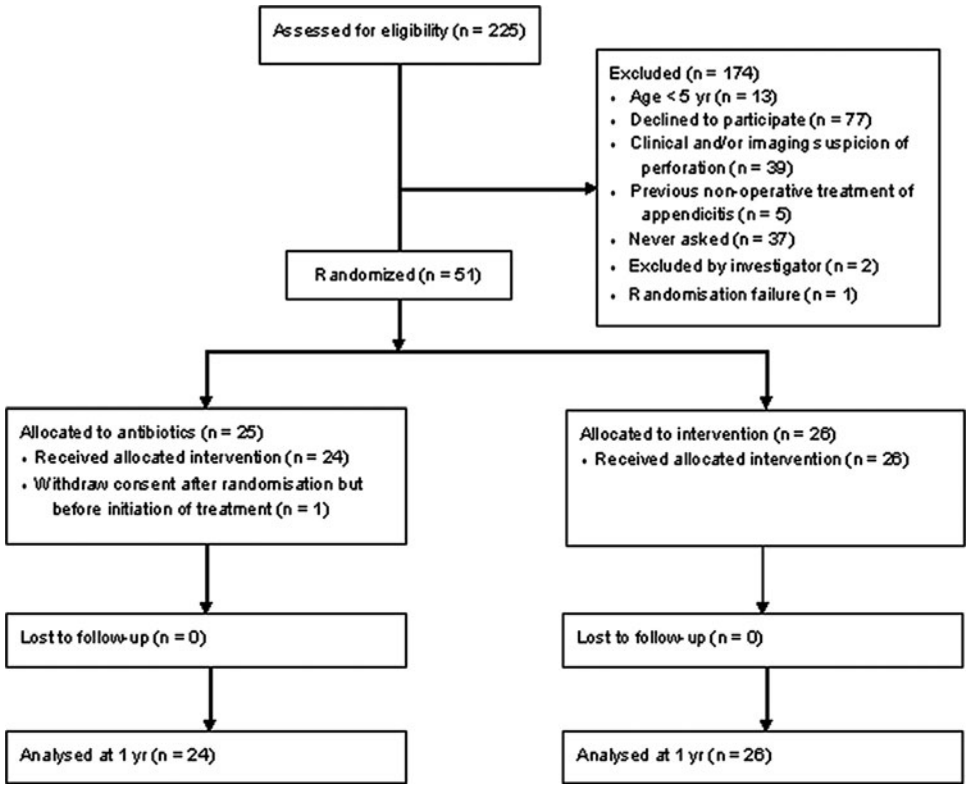


FIGURE 1. Study flowchart.

TABLE 1. Comparison of Participants, Those Eligible but Not Enrolled and Those Not Invited to Participate

	Randomized Children (n = 50)	Declined to Participate (n = 77)	*P**	Not Invited to Participate (n = 37)	*P*†
Age, yr	11.2 (5.9–15.0)	11.0 (5.8–14.9)	0.369	10.8 (5.3–14.9)	0.268
Male sex, n (%)	26 (52)	42 (55)	0.779	23 (62)	0.345
Duration of symptoms <48 h, n (%)	43 (86)	61 (79)	0.332	25 (68)	0.04
CRP at admission, mg/L	28 (1–185)	19 (1–152)	0.414	17.5 (1.0–150.0)	0.909
WBC ($\times 10^9$/L) at admission	14.3 (4.5–26.9)	15.0 (5.2–27.2)	0.086	15.0 (6.1–33.5)	0.297
Neutrophils ($\times 10^9$/L) at admission	11.5 (2.5–23.5)	12.5 (1.5–24.0)	0.155	3.6 (12.5–30.1)	0.295
Temperature at admission, °C	37.4 (36.3–39.0)	37.3 (35.9–37.3)	0.177	37.1 (35.7–39.3)	0.392

Data are median (range) unless specified.
*Comparison between randomized children and those who declined to participate.
†Comparison between randomized children and those who were not invited to participate.
CRP indicates C-reactive protein; WBC, white blood cells.

need to perform an appendectomy in children with longer symptom duration. The surgery and nonoperative treatment groups had similar demographic and admission characteristics (Table 2). All patients had at least 1 ultrasound examination, 1 had a second ultrasound scan, and 4 had a CT scan after the initial ultrasound scan. The reason for repeated examination was, in all cases, that the appendix was not seen at the initial examination.

Primary Outcome

All children randomized to surgery had a laparoscopic appendectomy with a 3-port technique. Histological examination confirmed the diagnosis of acute appendicitis in all cases (ie, no negative appendectomy, 21 phlegmonous appendicitis, 3 gangrenous appendicitis, and 2 perforated appendicitis), and there were no significant complications in this group.

All children randomized to nonoperative treatment with antibiotics received antibiotics per protocol. Two of these children had a significant complication. One child underwent an early appendectomy on day 2, as symptoms had failed to improve; a macroscopically normal appendix was removed and the child had a diagnosis of mesenteric lymphadenitis. Histological examination of the appendix was normal. This patient had had an inconclusive ultrasound scan and a CT scan suggestive of appendicitis with a tubular structure measuring 9 to 10 mm. The final report on this CT scan (produced after the surgery) was changed to a negative investigation. A second child returned to the emergency department on day 9 after randomization

TABLE 2. Comparison of Treatment Groups

	Randomized Children		
	Surgery (n = 26)	Nonoperative Treatment (n = 24)	*P*
Age, yr	11.1 (6.2–14.8)	12.2 (5.9–15.0)	0.130
Male sex, n (%)	12 (46)	14 (58)	0.389
Duration of symptoms <48 h, n (%)	23 (88)	20 (83)	0.602
CRP at admission, mg/L	27.0 (1.0–175.0)	30.5 (1.0–185.0)	0.892
WBC ($\times 10^9$/L) at admission	14.5 (4.5–26.9)	14.0 (4.8–19.0)	0.918
Neutrophils ($\times 10^9$/L) at admission	11.6 (2.9–23.5)	11.5 (2.5–16.8)	1.0
Temperature at admission, °C	37.5 (36.5–38.5)	37.3 (36.6–39.0)	0.199

Data are median (range) unless specified.
CRP indicates C-reactive protein; WBC, white blood cells.

with moderate abdominal pain after initial successful antibiotic treatment according to the study protocol. Ultrasound scan revealed signs of ongoing inflammation, and a walled-off perforated appendicitis was found at laparoscopic appendectomy. The primary outcome was similar in each group [appendectomy group 26/26 (100%) vs nonoperative treatment group 22/24 (92%); $P = 0.23$].

Secondary Outcomes

During the 1-year follow-up period, there were no significant or minor complications in the surgery group. In the nonoperative treatment group, there were no minor complications. However, 1 child had appendectomy for histopathologically confirmed recurrent acute appendicitis 9 months after randomization and 1 asymptomatic child underwent (histopathologically normal) appendectomy at parental request. A further 5 children returned with mild abdominal pain and had laparoscopic appendectomies at surgeon and parental discretion. All had a varying degree of fibrosis in the appendix but no inflammation. In all cases, symptoms resolved after surgery. Therefore, after 1-year of follow up, 15 of 24 children (62%) randomized to primary antibiotic treatment had not undergone an appendectomy.

Twelve children had a diagnosis of an appendicolith on imaging, 7 of 26 in the surgery group and 5 of 24 in the nonoperative treatment group ($P = 0.74$). Of the 5 children with an appendicolith in the nonoperative treatment group, 3 had appendectomy (none as primary failures, 1 due to recurrent acute appendicitis, 1 due to recurrent symptoms without appendicitis, and 1 on parental request). Thus, in the nonoperative treatment group, 2 children with an appendicolith did not have appendectomy within 1 year of follow-up, and of the total 9 who have had an appendectomy, only 3 had an appendicolith on imaging at the initial presentation.

Time from randomization to actual discharge home was calculated for each participant. The median time to discharge was significantly shorter in the surgical group [34.5 (16.2–95.0) hours] than in the nonoperative treatment group [51.5 (29.9–86.1) hours] ($P = 0.0004$). Despite this, the cost for the initial inpatient stay was significantly lower for the nonoperative treatment group [30,732 (18,980–63,863) SEK] than for the surgery group [45,805 (33,042–94,638) SEK] ($P < 0.0001$).

The total cost of treatment, including the cost of those patients having an appendectomy during the follow-up period, was similar in both treatment groups [nonoperative treatment 34,587 (19,120–146,552) SEK vs surgery 45,805 (33,042–94,638) SEK] ($P = 0.11$).

DISCUSSION

In this pilot RCT comparing nonoperative treatment with antibiotics and surgery for acute nonperforated appendicitis in children, we have shown that nonoperative treatment is feasible and safe. Overall, 40% of families asked to participate accepted and were enrolled, suggesting that nonoperative treatment is of interest to this patient population and their families. We consider it possible that in future randomized trials in children, this consent rate might be improved, as during the study we were unable to provide the parents with any evidence of safety or efficacy of antibiotics alone whereas future studies would have such evidence from this pilot trial. On the basis of the recruitment rate achieved, we believe a future RCT would be feasible.

Although this pilot trial was not adequately powered to detect differences in treatment efficacy, outcome data are useful to inform future studies. As defined, effective treatment was achieved in 100% and 92% in the surgery and nonoperative treatment groups, respectively. In the nonoperative treatment group, only 2 of 24 patients failed to meet criteria for the primary endpoint. One of them had mesenteric lymphadenitis, which may explain the failure to respond to antibiotics, as this patient's condition did not improve by antibiotic treatment. The other returned after initial resolution in symptoms with antibiotics and was found to have perforated appendicitis.

An important consideration for surgeons and parents after successful nonoperative treatment of acute appendicitis is the fate of the appendix. In this study, we did not offer routine interval appendectomy. A potential benefit of nonoperative treatment is the avoidance of an appendectomy (and associated general anesthesia) at all. For this benefit to be realized, the recurrent appendicitis rate must be low and acceptable to both surgeons and parents. In this study, there was one case of histologically proven recurrent appendicitis during the follow-up period (5%). However, a further 6 children had appendectomy within the 1-year follow-up period for reasons other than recurrent acute appendicitis including one at parental request. As this was a pilot trial of a novel treatment strategy (antibiotics for acute appendicitis in children), we were liberal with regard to indications for surgery during the follow-up period among children in the nonoperative treatment group. It is possible that patients in this group would not have had surgery if they had presented with their symptoms outside the trial setting. This may have contributed to the high rate of surgery during follow-up and raises the important question of what is an appropriate threshold for appendectomy in children who have been successfully discharged home after nonoperative treatment.

For nonoperative treatment to be considered equivalent to appendectomy, some may believe that the length of hospitalization should be similar. In this pilot trial, the postrandomization length of stay was longer for children in the nonoperative treatment group than for children undergoing appendectomy. A possible explanation for this is that we stipulated a minimum of 48 hours of intravenous antibiotics in our protocol. In the future, it may be possible to reduce this duration without affecting efficacy. During analysis of these time-related data, it became apparent that significant delays

between randomization and surgery will impact on the time from randomization to discharge and therefore potentially influence the interpretation of this outcome measure. Delays between randomization and surgery may occur due to hospital workload and/or time of presentation, as typically appendectomy is no longer performed during the night. Median time between randomization and surgery in this study was 5.8 hours but with a range of 0.8 to 26.2 hours. These factors must be considered carefully in any future RCT.

Although overall cost was similar between the 2 treatment groups, the cost of the initial inpatient treatment was significantly higher in the surgery group. Thus, the additional admissions for recurrent symptoms in the nonoperative treatment groups were a significant determinant of cost in this group. A cost-effectiveness analysis should be performed as part of any future study.

Although the number of patients treated nonoperatively was small, there were no safety issues either during the acute admission or during the follow-up period and so this trial provides no evidence that nonoperative treatment of acute appendicitis is unsafe. As this was a pilot trial with a relatively small sample size, the efficacy data produced should be interpreted with caution. Importantly, we do not recommend nonoperative treatment of simple acute appendicitis in all children until further large-scale efficacy studies have been completed. This pilot trial suggests that nonperforated acute appendicitis in children may be safely treated with antibiotics and that it would be appropriate and feasible to proceed to a similar larger, RCT to determine the efficacy of nonoperative treatment in this population.

REFERENCES

1. Svensson JF, Hall NJ, Eaton S, et al. A review of conservative treatment of acute appendicitis. *Eur J Pediatr Surg.* 2012;22:185–194.
2. Wilms IM, de Hoog DE, de Visser DC, et al. Appendectomy versus antibiotic treatment for acute appendicitis. *Cochrane Database Syst Rev.* 2011;9:CD008359.
3. Blakely ML, Williams R, Dassinger MS, et al. Early vs interval appendectomy for children with perforated appendicitis. *Arch Surg.* 2011;146:660–665.
4. Holcomb GW, III, St Peter SD. Current management of complicated appendicitis in children. *Eur J Pediatr Surg.* 2012;22:207–212.
5. Abes M, Petik B, Kazil S. Nonoperative treatment of acute appendicitis in children. *J Pediatr Surg.* 2007;42:1439–1442.
6. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c332.
7. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c869.

Randomized clinical trial

BJS Open Open Access

Randomized clinical trial of high *versus* low inferior mesenteric artery ligation during anterior resection for rectal cancer

S. Fujii[1,2], A. Ishibe[2], M. Ota[2], K. Watanabe[3], J. Watanabe[3], C. Kunisaki[2] and I. Endo[3]

[1]Department of Gastroenterological Surgery, Ichikawa Hospital, International University of Health and Welfare, Chiba, and [2]Department of Surgery, Gastroenterological Centre, and [3]Department of Gastroenterological Surgery, Graduate School of Medicine, Yokohama City University, Yokohama, Japan

Correspondence to: Professor S. Fujii, Department of Gastroenterological Surgery, Ichikawa Hospital, International University of Health and Welfare, 6-1-14 Kounodai, Ichikawa, Chiba, 232-0024, Japan (e-mail: sfujii@iuhw.ac.jp)

Background: **The optimal level for inferior mesenteric artery ligation during anterior resection for rectal cancer is controversial. The aim of this randomized trial was to clarify whether the inferior mesenteric artery should be tied at the origin (high tie) or distal to the left colic artery (low tie).**

Methods: **Patients were allocated randomly to undergo either high- or low-tie ligation and were stratified by surgical approach (open or laparoscopic). The primary outcome was the incidence of anastomotic leakage. Secondary outcomes were duration of surgery, blood loss and 5-year overall survival.**

Results: **Some 331 patients entered the trial between June 2006 and September 2012. The trial was stopped prematurely as recruitment was slow. Seven patients were excluded after randomization but before operation because of procedural changes. High tie and low tie were performed in 164 and 160 patients respectively. The incidence of anastomotic leakage was not significantly different (17·7 *versus* 16·3 per cent respectively; $P = 0·731$). The incidence of severe complications requiring intervention was 2·4 *versus* 5·0 per cent for high and low tie respectively ($P = 0·222$). In multivariable analysis, risk factors for anastomotic leakage included male sex (odds ratio 4·36, 95 per cent c.i. 1·56 to 12·18) and distance of the tumour from the anal verge (odds ratio 0·99, 0·98 to 1·00). At 5 years there were no significant differences in overall (87·2 *versus* 89·4 per cent respectively; $P = 0·386$) and disease-free (76·3 *versus* 77·6 per cent; $P = 0·765$) survival.**

Conclusion: **The level of ligation of the inferior mesenteric artery does not significantly influence the rate of anastomotic leakage. Registration number: NCT01861678 (https://clinicaltrials.gov).**

Funding information:
No funding information provided

Presented in part to the 68th General Meeting of the Japanese Society of Gastroenterological Surgery, Miyazaki, Japan, July 2013

Paper accepted 16 March 2018
Published online 8 June 2018 in Wiley Online Library (www.bjsopen.com). **DOI:** 10.1002/bjs5.71

Introduction

In rectal cancer surgery the inferior mesenteric artery (IMA) can be ligated at its origin from the aorta (high tie) or distal to the branch of the left colic artery (LCA) (low tie). High-tie ligation has been advocated[1–6] because it allows more radical resection and more accurate pathological staging. Others[7–13] favour low-tie ligation because of increased blood flow to the proximal end of the anastomosis. This debate goes all the way back to the descriptions by Miles[14] and Moynihan[15] in 1908. Recent studies[7–10] have recommended low tie, as there was no significant difference in survival rates between high- and low-tie ligation.

Some[16] have suggested that high tie should be restricted to patients with clinical suspicion of involved nodes around the origin of the IMA or to those who require additional vascular mobilization to construct a tension-free anastomosis. Japanese guidelines[17] recommend that upward lymph node dissection should be performed at the level of the IMA for clinical T2 or more advanced disease. There is no consensus, however, on where to divide the IMA. Several reviews[18–21] found no significant difference between high and low tie with regard to short- and long-term results, with different authors recommending different methods. All emphasized the need for RCTs[18,20,21]. In the present

study, patients with rectal cancer were randomized between high- and low-tie ligation.

Methods

This was a single-centre phase III RCT, conducted at Yokohama City University Medical Centre. About 80 patients with rectal cancer were operated on annually at this institute. Patients with rectal cancer who were scheduled to undergo anterior resection were eligible for inclusion. All tumours were defined according to the seventh edition of the Japanese General Rules for Clinical and Pathological Studies on Cancer of the Colon, Rectum and Anus[22]. The rectum is defined as the intestine between the level of the sacral promontory and the upper edge of the puborectal muscle. The clinical TNM classification for the staging of rectal cancer was based on colonoscopy, CT of the thorax, abdomen and pelvis, abdominal ultrasonography or MRI. The general condition of all patients undergoing elective surgery was assessed before surgery by an anaesthetist.

Inclusion criteria were: age 20 years or above and histologically proven adenocarcinoma of the rectum. Exclusion criteria were: a primary tumour that directly invaded another organ clinically (T4b), synchronous distant or peritoneal metastasis, operation scheduled as an emergency, previous history of colorectal surgery except for appendicectomy, active or recent treatment for malignancy in another organ, and multiple colorectal cancers that needed construction of two or more anastomoses. Pregnant and lactating women were excluded, as were patients scheduled for resection without colorectal anastomosis.

Patients provided written informed consent. The trial protocol was approved by the ethics committee of Yokohama City University. The trial was registered at https://clinicaltrials.gov (trial number NCT01861678).

Patients were allocated randomly to undergo high- or low-tie ligation of the IMA in a 1 : 1 ratio. Immediately before the operation, the surgeon in charge reported a registration to the Epidemiology Data Centre in the Department of Biostatistics, Yokohama City University, via the internal line of the hospital; randomization was done by an epidemiologist using the minimization method. To balance surgical backgrounds between high- and low-tie groups, patients were stratified by surgical approach (open or laparoscopic). The study was conducted by the open-label method. Blinding was not attempted.

Interventions

All surgical procedures were performed by a specialized colorectal treatment team. The surgeon in charge of the team had acquired a specialist qualification from the Japan Society of Coloproctology[23], which recognized his years of clinical experience in approved facilities and successful completion of the specialist qualifying examination[24]. Laparoscopic operations were performed by a surgeon who was similarly accredited by the Japanese Society for Endoscopic Surgery[25]. All operations were performed according to the standard procedure described in the seventh edition of the Japanese General Rules for Clinical and Pathological Studies on Cancer of the Colon, Rectum and Anus[22].

For high-tie ligation, the IMA was divided at its origin from the abdominal aorta. For low-tie ligation, the IMA was divided just after branching to the LCA. Dissection of lymph nodes around the IMA was added in low-tie ligation.

Conventional open surgery was performed in patients with bulky tumours (6 cm or larger). Other patients underwent laparoscopic surgery via a medial-to-lateral approach. The IMA was divided at the level according to allocation. Mobilization of the left colon was performed. The rectum distal to the tumour was divided with a linear stapler after rectal irrigation. Partial mesorectal excision was usually performed. The proximal colon was divided at least 10 cm from the lesion[22]. The distal margin was 3 cm for tumours above the peritoneal reflection and 2 cm for those in the mid and distal rectum[22].

A haemorrhage test of the marginal artery was performed at the planned side of division. When the artery did not bleed, the colon was resected until bleeding was confirmed.

Reconstruction was undertaken using an end-to-end double stapling technique. An air leak test was done after reconstruction. For laparoscopic operations, an abdominal incision longer than 8 cm was considered a conversion.

In open surgery, the left colon was dissected from the retroperitoneum and mobilized. Lymph node dissection around the IMA was performed. A pelvic side-wall lymphadenectomy was done for cT3–4 lower rectal cancers. As laparoscopic pelvic side-wall lymphadenectomy was not conducted at the start of the study, this procedure was performed by an open approach. All other steps were as described for the laparoscopic approach. Creation of a diverting stoma was left to the surgeon in charge. An intraluminal drainage tube was inserted from the anus in the absence of a diverting stoma.

The time from skin incision to ligation of the IMA was recorded. Complications were graded according to the Clavien–Dindo classification[26]. Complications occurring within 30 days of surgery were considered as early, and those beyond this time as late.

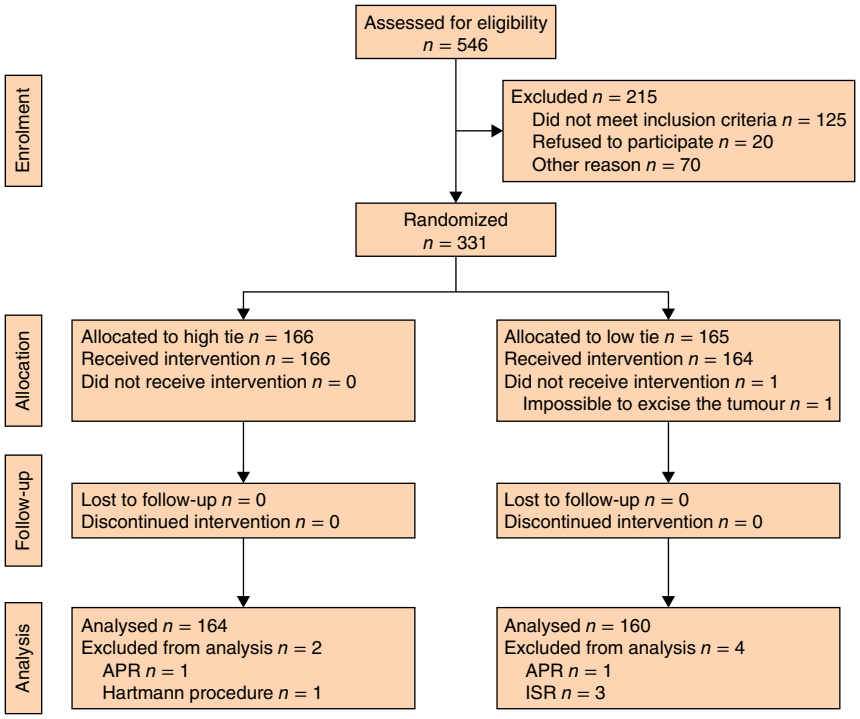


Fig. 1 CONSORT diagram for the trial. For high-tie ligation the inferior mesenteric artery was divided at its origin from the abdominal aorta; for low-tie ligation the inferior mesenteric artery was divided just after branching to the left colic artery. APR, abdominoperineal resection of rectum; ISR, intersphincteric resection of rectum

Pathology

Pathological results were recorded according to the seventh edition of the Japanese General Rules for Clinical and Pathological Studies on Cancer of the Colon, Rectum and Anus[22] and the seventh edition for TNM classification[27]. Total numbers of harvested lymph nodes were counted, as well as for each lymph node station separately. Lymph node stations were divided as: the area of IMA origin; the intermediate region along the IMA; and the perirectal region around the marginal vessels. Pathological proximal and distal margins were recorded, and circumferential margin involvement was defined as exposure of a cancer cell at the dissection surface on histological examination.

Neoadjuvant and adjuvant therapy, and follow-up

Patients did not have neoadjuvant chemotherapy, radiotherapy or chemoradiotherapy. When the pathological stage was IIb, IIc or III by histological examination, adjuvant treatment with oral or intravenous fluoropyrimidine-based chemotherapy was recommended.

The follow-up schedule was based on tumour stage. For stages 0 (defined as Tis in the Japanese Classification of Colorectal Carcinoma; adenocarcinoma was detected at the mucosa layer on the histological examination in stage 0) and I, follow-up included outpatient examinations with assessment of serum carcinoembryonic antigen (CEA), and chest, abdominal and pelvic CT once a year for 5 years. For stages II and IIIa, follow-up included outpatient examinations with assessment of serum CEA, and chest, abdominal and pelvic CT every 6 months for the first 3 years and once-yearly thereafter until 5 years after surgery. For stages IIIb and IIIc, follow-up included outpatient examinations with assessment of serum CEA, and chest, abdominal and pelvic CT every 4 months for the first 2 years, and every 6 months thereafter until 5 years after surgery. For stage IV disease, which was seen occasionally after randomization, the follow-up schedule was decided according to the condition of each patient.

Primary and secondary outcome measures

Primary outcome was the rate of anastomotic leakage. Leakage was defined as an incontinuity at the anastomosis detected clinically or radiologically. Contrast radiography via the drainage tube was not done in all patients. However, contrast radiography was performed in patients with purulent discharge via an abdominal drainage tube,

Table 1 Patient and tumour characteristics

	High tie ($n = 164$)	Low tie ($n = 160$)
Age (years)*	65·9(10·4)	65·6(11·5)
Sex ratio (M : F)	103 : 61	97 : 63
ASA grade		
1	39 (23·8)	53 (33·1)
2	115 (70·1)	95 (59·4)
3	10 (6·1)	12 (7·5)
ECOG performance status		
0	72 (43·9)	78 (48·8)
1	77 (47·0)	61 (38·1)
2	15 (9·1)	21 (13·1)
Prognostic Nutrition Index*	52·3(6·9)	52·2(5·3)
Concomitant disease†	116 (70·7)	102 (63·8)
Second synchronous colonic cancer	15 (9·1)	18 (11·3)
Cardiovascular disease	79 (48·2)	71 (44·4)
Diabetes	19 (11·6)	30 (18·8)
Other	57 (34·8)	48 (30·0)
History of laparotomy	21 (12·8)	28 (17·5)
BMI (kg/m^2)*	23·0(3·2)	22·4(3·5)
Tumour location		
Upper rectum	107 (65·2)	99 (61·9)
Lower rectum	57 (34·8)	61 (38·1)
Distance from anal verge (mm)*	88·7(32·9)	89·6(37·0)
Tumour diameter (mm)*	41·5(20·8)	41·9(20·5)
Histology		
Papillary adenocarcinoma	1 (0·6)	1 (0·6)
Well differentiated adenocarcinoma	81 (49·4)	85 (53·1)
Moderately differentiated adenocarcinoma	77 (47·0)	65 (40·6)
Poorly differentiated adenocarcinoma	1 (0·6)	2 (1·3)
Mucinous adenocarcinoma	3 (1·8)	2 (1·3)
Carcinoid tumour	1 (0·6)	4 (2·5)
Small cell carcinoma	0 (0)	1 (0·6)
pTNM stage		
0	4 (2·4)	6 (3·8)
1	56 (34·1)	54 (33·8)
2	43 (26·2)	36 (22·5)
3	54 (32·9)	56 (35·0)
4	7 (4·3)	8 (5·0)
Surgical approach		
Open	57 (34·8)	52 (32·5)
Laparoscopic	107 (65·2)	108 (67·5)
Level of anastomosis from anal verge (cm)*	5·8(2·0)	5·7(2·1)
Diverting stoma	36 (22·0)	47 (29·4)
Insertion of intraluminal drain from anus	12 (7·3)	19 (11·9)
Simultaneous resection of other organ	6 (3·7)	10 (6·3)
No. of linear stapler cartridges used		
1	120 (73·2)	111 (69·4)
≥ 2	44 (26·8)	49 (30·6)
Pelvic side-wall lymphadenectomy	25 (15·2)	22 (13·8)
Adjuvant chemotherapy	39 (23·8)	46 (28·8)

Values in parentheses are percentages unless indicated otherwise; *values are mean(s.d.). †Some patients had more than one concomitant disease. ECOG, Eastern Cooperative Oncology Group.

or peritonitis. If fistula was confirmed by contrast radiography, a patient was diagnosed as having an anastomotic leakage. Anastomotic leakage was categorized according to the Clavien–Dindo system[26].

Secondary outcomes were duration of surgery, blood loss and 5-year overall survival rate.

Table 2 Short-term outcomes

	High tie ($n = 164$)	Low tie ($n = 160$)	P†
Anastomotic leakage			
All grades	29 (17·7)	26 (16·3)	0·731
Grade 2 or above	16 (9·8)	14 (8·8)	0·755
Grade 3 or above	4 (2·4)	8 (5·0)	0·222
Leakage grade			
1	13 (7·9)	12 (7·5)	0·432
2	12 (7·3)	6 (3·8)	
3	4 (2·4)	7 (4·4)	
4	0 (0)	0 (0)	
5	0 (0)	1 (0·6)	
Mortality	0 (0)	1 (0·6)	0·311
Early complication (except leakage)	61 (37·2)	56 (35·0)	0·681
Surgical-site infection	8 (4·9)	10 (6·3)	0·590
Ileus	16 (9·8)	8 (5·0)	0·102
Enteritis	2 (1·2)	2 (1·3)	1·000
Chylous ascites	3 (1·8)	5 (3·1)	0·452
Urinary tract infection	1 (0·6)	2 (1·3)	0·547
Urinary dysfunction	4 (2·4)	3 (1·9)	0·727
Conversion to open surgery	6 of 107 (5·6)	2 of 108 (1·9)	0·142
Estimated blood loss (ml)*	155(299)	152(289)	0·867‡
Blood transfusion	3 (1·8)	3 (1·9)	0·976
Duration of surgery (min)*	209(67)	206(59)	0·672‡
Duration of IMA tie from start (min)*	41(15)	52(15)	< 0·001‡
Duration of laparoscopic procedure (min)*	161(42)	165(45)	0·525‡
Postoperative hospital stay (days)*	17(14)	16(12)	0·451‡

Values in parentheses are percentages unless indicated otherwise; *values are mean(s.d.). IMA, inferior mesenteric artery. †χ^2 test, except ‡Student's t test.

Statistical analysis

It was hypothesized that low-tie ligation would decrease the rate of anastomotic leakage from 15 to 6 per cent. Using a power of 80 per cent and α of 0·05, a sample size of 362 patients was needed. A dropout rate of approximately 10 per cent was anticipated. Therefore, 400 patients had to be included in this study. Enrolment was scheduled for 5 years after inclusion of the first patient.

Data were analysed according to the intention-to-treat principle. SAS® software version 9.2 for Windows® (SAS Institute, Cary, North Carolina, USA) was used for statistical analysis.

Categorical variables are presented as frequencies and percentages. Continuous variables are presented as mean(s.d.) values. The χ^2 test and Student's t test were used to compare categorical and continuous variables respectively. Survival was analysed by the Kaplan–Meier method, and the difference between high- and low-tie ligation was analysed with the log rank test. Risk factors for anastomotic leakage were assessed by logistic regression using a forward method. Variables with $P < 0·100$ were entered

Table 3 Risk factors for anastomotic leakage in all grades

	Leakage		Univariable analysis		Multivariable analysis	
	Yes ($n = 55$)	No ($n = 269$)	Odds ratio†	*P*	Odds ratio†	*P*
Sex						
M	47 (23·5)	153 (76·5)	4·40 (2·00, 9·67)	< 0·001	4·36 (1·56, 12·18)	0·005
F	8 (6·5)	116 (93·5)	1·00 (reference)		1·00 (reference)	
pT category						
pTis–pT3	39 (14·8)	225 (85·2)	1·00 (reference)		1·00 (reference)	
pT4	16 (27)	44 (73)	2·11 (1·08, 4·11)	0·028	1·57 (0·58, 4·24)	0·371
Blood transfusion						
Yes	4 (67)	2 (33)	10·28 (1·83, 57·59)		1·27 (0·06, 25·97)	0·875
No	51 (16·0)	267 (84·0)	1·00 (reference)	0·008	1·00 (reference)	
Conversion of laparoscopy						
Yes	4 (50)	4 (50)	5·09 (1·21, 21·34)		3·14 (0·69, 14·25)	0·139
No	34 (16·4)	173 (83·6)	1·00 (reference)	0·026	1·00 (reference)	
Distance of tumour from anal verge (mm)*	81(31)	91(35)	0·99 (0·98, 1·00)	0·043	0·99 (0·98, 1·00)	0·011
No. of stapler firings*	1·30(0·56)	1·49(0·64)	1·64 (1·05, 2·56)	0·031	1·09 (0·62, 1·91)	0·765

Values in parentheses are percentages unless indicated otherwise; *values are mean(s.d.) and †95 per cent confidence intervals in parentheses.

into multivariable analysis. $P < 0·050$ was considered statistically significant. All analyses were two-sided.

Results

Some 331 patients were randomized between June 2006 to September 2012 (*Fig. 1*). Due to slow recruitment, the trial was stopped prematurely. One hundred and sixty-six patients were assigned to the high-tie group and 165 to the low-tie group. Two patients in the high-tie group were excluded because of changes in operative procedure: one underwent an abdominoperineal rectal resection (APR) and the other had a Hartmann procedure. Five patients in the low-tie group were excluded because of changes in operative procedure: three patients underwent an intersphincteric rectal resection, one had an APR, and one tumour could not be excised. In the low-tie group, the LCA of one patient was separated during operation because of a high-tension anastomosis. This patient was analysed in the low-tie group according to allocation.

The primary outcome could be analysed for 164 patients in the high-tie group and 160 in the low-tie group. Patient and tumour characteristics are shown in *Table 1*. Second synchronous colorectal carcinomas were seen fairly frequently and mainly included Tis tumours within 10 cm of the rectal carcinoma necessitating additional distal sigmoid resections. Simultaneous resection of another organ was needed by six patients in the high-tie group (hysterectomy, 3; oophorectomy, 1; partial resection of the jejunum, 2) and ten in the low-tie group (hysterectomy, 4; oophorectomy, 2; partial resection of the ileum, 2; partial resection of the urinary bladder, 2).

Anastomotic leakage

The overall rate of anastomotic leakage was 17·7 per cent in the high-tie group and 16·3 per cent in the low-tie group ($P = 0·731$) (*Table 2*). For grade 2 or higher leakage the rate was 9·8 and 8·8 per cent respectively ($P = 0·755$), and for grade 3 or higher it was 2·4 and 5·0 per cent ($P = 0·222$). The 12 patients with grade 3 and 5 anastomotic leakage underwent reoperation with creation of a stoma. All but the one patient who died underwent stoma reversal later.

Risk factors for anastomotic leakage

In univariable analysis, male sex, advanced T status (T4), transfusion of red blood cells, conversion to open surgery, distance of tumour from the anal verge and the number of stapler cartridges fired were associated significantly with the rate of anastomotic leakage. In multivariable analysis, male sex and distance of tumour from the anal verge were identified as independent risk factors for anastomotic leakage (*Table 3*).

Surgical parameters, complications and pathology

Duration of surgery did not differ significantly between groups, although time to ligation of the IMA was significantly longer in the low-tie group ($P < 0·001$) (*Table 2*). Blood loss did not differ significantly between the groups. The overall early complication rate was not significantly different, and neither was the total number of lymph nodes harvested or the number of lymph nodes per station. Proximal and distal pathological margins, and the positive circumferential margin rate were similar in high- and low-tie

Table 4 Oncological quality of surgery

	High tie ($n = 164$)	Low tie ($n = 160$)	P†
No. of lymph nodes harvested*			
Total	26·4(11·4)	24·1(12·2)	0·079‡
IMA root nodes	2·8(2·1)	2·9(2·7)	0·639‡
Intermediate lymph nodes	5·4(3·9)	5·1(3·9)	0·623‡
Perirectal lymph nodes	15·5(7·6)	14·1(7·5)	0·130‡
Lymph node involvement at each station			
IMA root nodes	3 (1·8)	5 (3·1)	0·452
Intermediate lymph nodes	10 (6·1)	8 (5·0)	0·666
Perirectal lymph nodes	58 (35·4)	55 (34·4)	0·852
IMA root nodes positive, intermediate nodes negative	0 (0)	1 (0·6)	0·311
Intermediate nodes positive, perirectal nodes negative	1 (0·6)	1 (0·6)	0·986
Pathological proximal margin (cm)*	13·4(5·5)	12·5(4·9)	0·110‡
Pathological distal margin (cm)*	3·1(1·8)	3·2(2·0)	0·618‡
Positive circumferential margin	3 (1·8)	5 (3·1)	0·452

*Values are mean(s.d.). IMA, inferior mesenteric artery. †χ^2 test, except ‡Student's t test.

Table 5 Long-term results

	High tie ($n = 164$)	Low tie ($n = 160$)	P*
5-year overall survival rate (%)			
All stages	87·2	89·4	0·386
Stage 1	94·2	96·3	0·740
Stage 2	87·8	84·6	0·965
Stage 3	88·2	88·6	0·880
Stage 4	28·6	72·9	0·109
5-year relapse-free survival rate (%)			
All stages	76·3	77·6	0·765
Stage 1	94·2	86·5	0·187
Stage 2	78·1	79·1	0·856
Stage 3	66·2	72·9	0·314

*Log rank test.

groups (*Table 4*). All patients with positive IMA root nodes also had positive intermediate or perirectal lymph nodes.

Long-term results

The 5-year overall survival rate did not differ significantly between high- and low-tie groups (87·2 *versus* 89·4 per cent respectively; $P = 0·386$). Neither did the 5-year relapse-free survival rate: 76·3 *versus* 77·6 per cent ($P = 0·765$). Significant differences regarding survival were not detected within stages (*Table 5*).

Discussion

In this study the level of IMA ligation in rectal cancer surgery did not affect the anastomotic leak rate. Other factors, including male sex and distance of the tumour from the anal verge, did have an independent influence on the risk of clinical anastomotic leakage. Although this study was stopped prematurely, it is unlikely that level of ligation of the IMA adds significantly to the risk of anastomotic leakage.

Several studies[11–13] have shown that colonic blood flow is decreased in high-tie compared with low-tie ligation. Two reports[28,29] have described the development of proximal bowel necrosis or ischaemia after high-tie ligation (2 and 0·8 per cent respectively). A haemorrhage test was performed in the present study, but blood flow was not evaluated in a quantitative manner. Differences were not observed between the groups for nearly all short-term results. The longer IMA tie time in the low-tie group probably reflects the technical complexity involved in preservation of the LCA. This did not significantly increase total operating time. Some studies[30,31] have reported that an IMA branching pattern with a large distance between the origins of the IMA and LCA causes technical difficulty. It would appear that surgeons should not hesitate to change the tie level of the IMA when performing a difficult low-tie ligation.

No significant differences in long-term results were detected between the two groups, suggesting that low tie with lymph node dissection around the IMA has validity as a surgical treatment. The numbers of lymph nodes harvested around the IMA root and total lymph nodes were not significantly different between groups, in agreement with previous reports[3,4]. Thus both approaches appear equal from an oncological perspective.

This single-centre study has several limitations. It was stopped prematurely because of slow accrual. This may have introduced bias, although it is unlikely that inclusion of the number of patients assumed in the power calculations would have changed the most important results as the differences between groups were small. Functional

evaluations, such as defaecation, digestive symptoms, bladder and sexual functions, were not performed. Lange and colleagues[19] recommended low-tie ligation because it allows preservation of the autonomous innervation of the proximal colon. Another study[32], however, found no difference in defaecatory function or postoperative complications in a relatively small randomized trial. Shiomi and co-workers[33] reported that the incidence of anastomotic leakage was lower for low than for high tie in a prospective multicentre cohort study. A randomized multicentre study[34] is currently in progress. Finally, patients and treatment schedules may differ between Japan and countries in the West, where a significant proportion of patients would have had neodjuvant radiotherapy or chemoradiotherapy, unlike patients in the present study.

Acknowledgements

The authors thank M. S. Oba for help with the statistical analysis.
Disclosure: The authors declare no conflict of interest.

References

1 Gabriel WB, Dukes CE, Bussey HJR. Lymphatic spread in cancer of the rectum. *Br J Surg* 1935; **23**: 395–413.
2 Morgan CN, Griffiths JD. High ligation of the inferior mesenteric artery during operations for carcinoma of the distal colon and rectum. *Surg Gynecol Obstet* 1959; **108**: 641–650.
3 Kanemitsu Y, Hirai T, Komori K, Kato T. Survival benefit of high ligation of the inferior mesenteric artery in sigmoid colon or rectal cancer surgery. *Br J Surg* 2006; **93**: 609–615.
4 Chin CC, Yeh CY, Tang R, Changchien CR, Huang WS, Wang JY. The oncologic benefit of high ligation of the inferior mesenteric artery in the surgical treatment of rectal or sigmoid colon cancer. *Int J Colorectal Dis* 2008; **23**: 783–788.
5 Alici A, Kement M, Gezen C, Akin T, Vural S, Okkabaz N *et al.* Apical lymph nodes at the root of the inferior mesenteric artery in distal colorectal cancer: an analysis of the risk of tumor involvement and the impact of high ligation on anastomotic integrity. *Tech Coloproctol* 2010; **14**: 1–8.
6 Kessler H, Hohenberger W. Extended lymphadenectomy in colon cancer is crucial. *World J Surg* 2013; **37**: 1789–1798.
7 Pezim ME, Nicholls RJ. Survival after high or low ligation of the inferior mesenteric artery during curative surgery for rectal cancer. *Ann Surg* 1984; **200**: 729–733.
8 Surtees P, Ritchie JK, Phillips RK. High *versus* low ligation of the inferior mesenteric artery in rectal cancer. *Br J Surg* 1990; **77**: 618–621.
9 Uehara K, Yamamoto S, Fujita S, Akasu T, Moriya Y. Impact of upward lymph node dissection on survival rates in advanced lower rectal carcinoma. *Dig Surg* 2007; **24**: 375–381.
10 Yasuda K, Kawai K, Ishihara S, Murono K, Otani K, Nishikawa T *et al.* Level of arterial ligation in sigmoid colon and rectal cancer surgery. *World J Surg Oncol* 2016; **14**: 99.
11 Seike K, Koda K, Saito N, Oda K, Kosugi C, Shimizu K *et al.* Laser Doppler assessment of the influence of division at the root of the inferior mesenteric artery on anastomotic blood flow in rectosigmoid cancer surgery. *Int J Colorectal Dis* 2007; **22**: 689–697.
12 Dworkin MJ, Allen-Mersh TG. Effect of inferior mesenteric artery ligation on blood flow in the marginal artery-dependent sigmoid colon. *J Am Coll Surg* 1996; **183**: 357–360.
13 Komen N, Slieker J, de Kort P, de Wilt JH, van der Harst E, Coene PP *et al.* High tie *versus* low tie in rectal surgery: comparison of anastomotic perfusion. *Int J Colorectal Dis* 2011; **26**: 1075–1078.
14 Miles WE. A method of performing abdomino-perineal excision for carcinoma of the rectum and of the terminal portion of the pelvic colon. *Lancet* 1908; **2**: 1812–1813.
15 Moynihan BG. The surgical treatment of cancer of the sigmoid flexure and rectum. *Surg Gynecol Obstet* 1908; **6**: 463–466.
16 Bleday R, Garcia-Aguilar J. Surgical treatment of rectal cancer. In *The ASCRS Textbook of Colon and Rectal Surgery*, Wolff BG, Fleshman JW, Beck DE, Pemberton JH, Wexner SD (eds). Springer: New York, 2007; 413–436.
17 Japanese Society for Cancer of the Colon and Rectum. *JSCCR Guidelines 2010 for the Treatment of Colorectal Cancer*. Kanehira-Syuppan: Tokyo, 2010.
18 Titu LV, Tweedle E, Rooney PS. High tie of the inferior mesenteric artery in curative surgery for left colonic and rectal cancers: a systematic review. *Dig Surg* 2008; **25**: 148–157.
19 Lange MM, Buunen M, van de Velde CJ, Lange JF. Level of arterial ligation in rectal cancer surgery: low tie preferred over high tie. A review. *Dis Colon Rectum* 2008; **51**: 1139–1145.
20 Hida J, Okuno K. High ligation of the inferior mesenteric artery in rectal cancer surgery. *Surg Today* 2013; **43**: 8–19.
21 Cirocchi R, Trastulli S, Farinella E, Desiderio J, Vettoretto N, Parisi A *et al.* High tie *versus* low tie of the inferior mesenteric artery in colorectal cancer: a RCT is needed. *Surg Oncol* 2012; **21**: e111–e123.
22 Japanese Society for Cancer of Colon and Rectum. [*General Rules for Clinical and Pathological Studies on Cancer of the Colon, Rectum and Anus*.] (7th edn). Kanehira-Syuppan: Tokyo, 2006.
23 Japan Society of Coloproctology. *Specialist's List* [*The Japan Society of Coloproctology web site*]; 2013. http://www.coloproctology.gr.jp/doctor/ [accessed 25 May 2017].
24 Japan Society of Coloproctology. *Specialist System Rule* [*The Japan Society of Coloproctology web site*]; 2009. http://www.coloproctology.gr.jp/pdf/200912specialistrules.pdf [accessed 25 May 2017].

25 Mori T, Kimura T, Kitajima M. Skill accreditation system for laparoscopic gastroenterologic surgeons in Japan. *Minim Invasive Ther Allied Technol* 2010; **19**: 18–23.
26 Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004; **240**: 205–213.
27 International Union Against Cancer. Colon and Rectum. In *TNM Classification of Malignant Tumours* (7th edn). Wiley–Liss: New York, 2010; 143–164.
28 Tsujinaka S, Kawamura YJ, Tan KY, Mizokami K, Sasaki J, Maeda T *et al.* Proximal bowel necrosis after high ligation of the inferior mesenteric artery in colorectal surgery. *Scand J Surg* 2012; **101**: 21–25.
29 Park MG, Hur H, Min BS, Lee KY, Kim NK. Colonic ischemia following surgery for sigmoid colon and rectal cancer: a study of 10 cases and a review of the literature. *Int J Colorectal Dis* 2012; **27**: 671–675.
30 Bertrand MM, Delmond L, Mazars R, Ripoche J, Macri F, Prudhomme M. Is low tie ligation truly reproducible in colorectal cancer surgery? Anatomical study of the inferior mesenteric artery division branches. *Surg Radiol Anat* 2014; **36**: 1057–1062.
31 Patroni A, Bonnet S, Bourillon C, Bruzzi M, Zinzindohoué F, Chevallier JM *et al.* Technical difficulties of left colic artery preservation during left colectomy for colon cancer. *Surg Radiol Anat* 2016; **38**: 477–484.
32 Matsuda K, Hotta T, Takifuji K, Yokoyama S, Oku Y, Watanabe T *et al.* Randomized clinical trial of defaecatory function after anterior resection for rectal cancer with high *versus* low ligation of the inferior mesenteric artery. *Br J Surg* 2015; **102**: 501–508.
33 Shiomi A, Ito M, Maeda K, Kinugasa Y, Ota M, Yamaue H *et al.* Effects of a diverting stoma on symptomatic anastomotic leakage after low anterior resection for rectal cancer: a propensity score matching analysis of 1014 consecutive patients. *J Am Coll Surg* 2015; **220**: 186–194.
34 Mari G, Maggioni D, Costanzi A, Miranda A, Rigamonti L, Crippa J *et al.* High or low inferior mesenteric artery ligation in laparoscopic low anterior resection: study protocol for a randomized controlled trial (HIGHLOW trial). *Trials* 2015; **16**: 21.

Original article

BJS Open Open Access

Cohort study of the effect of surgical repair of symptomatic diastasis recti abdominis on abdominal trunk function and quality of life

A. Olsson[1,3], O. Kiwanuka[1,3], S. Wilhelmsson[4], G. Sandblom[1,3] and O. Stackelberg[1,2,3]

[1]Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset, and [2]Unit of Cardiovascular and Nutritional Epidemiology Institute of Environmental Medicine, Karolinska Institutet, [3]Department of Surgery, Södersjukhuset, and [4]Functional Area Occupational Therapy and Physiotherapy, Allied Health Professionals Function, Karolinska University Hospital, Stockholm, Sweden

Correspondence to: Dr A. Olsson, Department of Clinical Science and Education and Department of Surgery, Södersjukhuset, Sjukhusbacken 10, 118 83 Stockholm, Sweden (e-mail: anders.olsson@ki.se)

Background: **During pregnancy, women are at risk of developing persistent symptomatic diastasis recti abdominis (DRA), which may have a detrimental effect on their physical function and quality of life (QoL). The aim of this prospective cohort study was to determine the effect of surgical repair of DRA on abdominal trunk function, urinary incontinence and QoL in postpartum women with trunk instability symptoms resistant to training.**

Methods: **Postpartum women with diagnosed DRA and training-resistant symptoms underwent double-row plication of the linea alba. Abdominal trunk function was evaluated as the primary endpoint using a multimodal examination tool, the Abdominal Trunk Function Protocol. Recurrence was assessed by CT, urinary incontinence was evaluated using the Urogenital Distress Inventory (UDI-6) and Incontinence Impact Questionnaire (IIQ-7), and QoL was assessed with the Short Form 36 (SF-36®) questionnaire. All subjects were examined before and 1 year after surgery.**

Results: **Sixty women were recruited. There was no DRA recurrence at the 1-year follow-up. Self-reported abdominal trunk function had improved in 98 per cent of patients, with a mean score improvement of 79·1 per cent. In the physiological tests monitored by a physiotherapist, 76 per cent performed better and endured exercise tests longer than before surgery. All SF-36® subscales improved significantly compared with preoperative scores and reached levels similar to, or higher than, the normative Swedish female population. For the UDI-6 and IIQ-7, 47 and 37 per cent respectively reported fewer symptoms at follow-up than before surgery, and 13 and 8 per cent respectively reported more symptoms.**

Conclusion: **In this series of postpartum women presenting with DRA and symptoms of trunk instability resistant to training, surgical reconstruction resulted in a significant improvement in abdominal trunk function, urinary incontinence and QoL.**

Funding information
Stockholm County Council
Swedish Society of Medicine

Paper accepted 26 June 2019
Published online 11 September 2019 in Wiley Online Library (www.bjsopen.com). **DOI:** 10.1002/bjs5.50213

Introduction

Diastasis recti abdominis (DRA) is a common and expected condition during pregnancy[1], owing to mechanical stretching, expansion and hormonal changes[2]. The condition is often characterized by bulging or sagging in the abdominal midline during abdominal muscle contraction. Although no consensus regarding the definition of DRA exists, it is often defined in the literature as separation of the recti greater than 22–30 mm[3–5].

DRA usually regresses to its prepregnancy width, but the condition persists in 32–46 per cent of postpartum women[6–9]. Reported risk factors for persistent DRA include maternal age, multiparity, caesarean section, macrosomia and multiple gestations[10]. A persistent DRA may be associated with abdominal trunk instability, which

could result in development of lower back pain, lack of trunk strength and urinary incontinence[11–13]. However, it remains unclear whether the DRA actually causes these symptoms or not. Although inconclusive[7,8,14], persistent lower back pain after pregnancy has been reported in 11–21 per cent of postpartum women[15–17].

The management of DRA is also a subject of discussion. Conservative management with training and weight loss is generally advised as first-line treatment. There is no strong evidence that training during pregnancy and in the postpartum period decreases the risk of persistent DRA[18–20], although some studies[8,21] have reported that specific exercises could increase abdominal trunk stability and reduce some of the associated symptoms. Surgical reconstruction has been reported to restore abdominal trunk function[22–26] and improve lower back pain and urinary incontinence[24,26]. General awareness of symptomatic DRA is poor, and patients are commonly advised to undertake non-specific physical training, or told that the condition is only cosmetic in nature. To evaluate potential treatments for persistent symptomatic DRA, a standardized and comprehensive multimodal protocol, able to capture the wide panorama of dysfunctions associated with the condition, is required.

Thus, symptomatic persistent DRA lacks clarity of definition and management[18,20]. This study aimed to evaluate the effect of surgical reconstruction of DRA in postpartum women, where no improvement in symptoms had been achieved by adequate physical training, using a standardized multimodal examination.

Methods

Women with symptomatic DRA were recruited between January 2015 and March 2017. All first underwent ultrasound measurement of the DRA, CT to localize any concomitant ventral hernia, and assessment of trunk function by a physiotherapist. Potential candidates received an individualized trunk stabilization training programme, and were re-evaluated by a physiotherapist after 3–6 months. Candidates presenting with subjective training-resistant trunk symptoms after evaluation were considered eligible for inclusion. Inclusion criteria were: non-smoker, age 18–55 years, BMI below 35 kg/m^2, DRA greater than 30 mm on ultrasound imaging at any level, trunk instability symptoms persisting after more than 3 months of standardized trunk stability training, more than 1 year from last delivery, and no intention of further pregnancy. The presence of preoperative cosmetic issues was not considered as a symptom or as an outcome in this study. Written informed consent was obtained from all participants before inclusion. The Regional Ethics Committee, Karolinska Institutet, Stockholm, approved the study. The local ethics committee approved all procedures (Dnr. 2015/1753-31).

Evaluation of symptoms

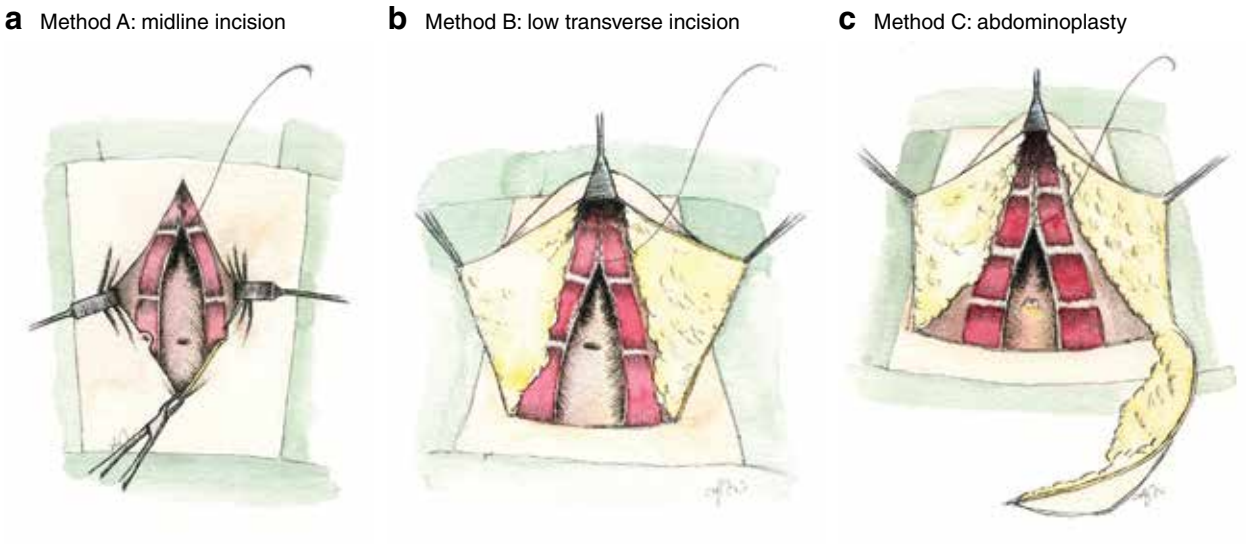
The primary outcome was abdominal trunk function. Secondary outcomes were quality of life (QoL), urinary incontinence and DRA recurrence (at follow-up).

To evaluate abdominal trunk function, a standardized multimodal trunk function test was designed to cover all dysfunctions associated with symptomatic DRA: the Abdominal Trunk Function Protocol (ATFP) (*Appendix S1*, supporting information). The ATFP consists of a self-rating section, where the participants score physical function (Disability Rating Index (DRI)), and seven trunk function tests supervised and monitored by a physiotherapist following a strict schema. The validated DRI covers 12 non-specific activities of daily life, each one self-rated on a visual analogue scale of 0–100 mm, providing a score of 0–100 for each activity, where 0 represents no difficulty in performing the specific task and 100 indicates an inability to perform the task at all[27]. The seven trunk function tests have been validated separately and measure different aspects of trunk and pelvic strength, endurance and stability. They are: the back muscle strength test, the abdominal muscle strength test, the lateral core stability test (left and right side), the ventral core stability test, the active straight-leg-raising test, and the pelvic joint provocation test[28,29]. The trunk function tests were conducted and monitored by a physiotherapist. The ATFP evaluation was performed before and 1 year after surgery. QoL was evaluated using the self-reported Medical Outcome Survey Short Form 36 (SF-36®) (Rand Corporation, Santa Monica, California, USA)[30]. Urinary incontinence was evaluated using the self-reported Urogenital Distress Inventory (UDI-6) and the Incontinence Impact Questionnaire (IIQ-7) forms[31]. DRA recurrence was assessed by CT 1 year after surgery, and was defined as a persisting diastasis greater than 30 mm.

Surgical reconstruction technique

The surgical procedure was a standardized suture repair of the diastasis using a double-row plication with absorbable Quill™ 2/0 sutures (Angiotech, Reading, Pennsylvania, USA). Access to the linea alba depended on anatomical conditions, body figure and excess skin. The surgical procedure was categorized according to the incision made: method A used a midline incision; method B involved a low transverse incision, including limited resection of excessive

Fig. 1 Illustrations of the three surgical procedures



Standardized double-layer plication of the linea alba was used without entering the rectus sheet, with absorbable self-retaining Quill™ 2/0 suture. **a** Method A: midline incision with a suture repair and without skin excision. **b** Method B: low transverse incision with a suture repair and limited skin excision with a floating umbilicus. **c** Method C: low transverse incision with a suture repair, skin excision and umbilical transposition.

skin and a floating umbilicus; and method C employed abdominoplasty, including resection of excessive skin and umbilical transposition (*Fig. 1a–c*). The three methods had an identical deep muscle layer technique. The decision of which method to use was based on the anatomical circumstances and the woman's preference (after detailed information regarding the risks associated with the different incisions).

All surgical procedures except one were performed by one of two consultant surgeons, at either the general surgical unit or the ambulatory surgical unit. All women were admitted to the ward for postoperative care. Patients operated on with methods B and C had an active 14-Fr catheter drain(s) that was removed when fluid loss was less than 50 ml/day according to local routines for ventral hernia repair. All women were instructed to wear a girdle for 12 weeks (day and night in weeks 1–8, daytime only in weeks 9–12), which has been standard (with minor adjustments) in previous studies[32]. Patients were also instructed to participate in a standardized rehabilitation programme developed by the physiotherapy department at the authors' hospital (*Appendix S2*, supporting information), as well as daily exercise such as short walks, but to avoid heavy physical exercise during the first 12 weeks. All women were followed up clinically and with ultrasound assessment at 6–8 weeks and 1 year after surgery. All postoperative complications within 30 days were registered.

Statistical analysis

Descriptive statistics were used to characterize demography. Pairwise correlation coefficients were performed between measurements of rectus diastasis, comparing preoperative ultrasound and CT scan measurements with the width measured at surgery. For continuous variables, paired *t* tests and Wilcoxon signed rank tests were used to identify changes in symptoms at 1-year follow-up, and McNemar's test was used to evaluate contingency of dichotomous variables. All tests were two-sided and considered statistically significant at a level of $P \leq 0{\cdot}050$. For the DRI, each parameter was investigated individually and the total DRI score was used for comparison purposes. SF-36® results were analysed and compared with data from 2679 women aged 15–44 years in the Swedish SF-36 Health Survey[30] (*Table S1*, supporting information). Linear regression was used to test whether the degrees of preoperative symptoms were associated with change in those symptoms after surgery. Non-linearity was investigated by adding a quadratic term of the preoperative variable investigated in the model. Statistical analyses were performed using Stata® 12.1 (StataCorp, College Station, Texas, USA).

Results

Table 1 summarizes the preoperative demographics of the 60 women who were included in the study. Their mean

Table 1 Preoperative characteristics of women who had surgery for diastasis recti abdominis

	No. of patients* (n = 60)
Age (years)†	38·8(5·5)
BMI (kg/m²)‡	22·6 (17·2–36·0)
No. of births‡	2 (1–5)
Vaginal delivery	2 (1–4)
Caesarean section	2 (1–4)
Time from last birth to surgery (months)‡	34 (12–192)
Duration of training before surgery (months)‡	7 (3–24)
Size of diastasis recti (cm)‡	
Ultrasonography	4·5 (3·0–9·0)
CT	5·0 (1·0–10·0)
Perioperative finding	4·5 (3·0–9·0)
Ventral hernia	45 (75)

*With percentages in parentheses unless indicated otherwise; values are †mean(s.d.) and ‡median (range).

age at the time of surgery was 38·8 (range 20·5–53) years. The follow-up rate was 93 per cent (56 of 60) for the DRI questionnaire, and 83 per cent (50 of 60) for the seven functional tests. There were four dropouts due to subacute orthopaedic surgery (1 patient), emigration (1), unrelated psychiatric disability (1) and declined further participation (1). A further six participants were excluded from follow-up of the functional tests owing to incomplete forms.

Surgery

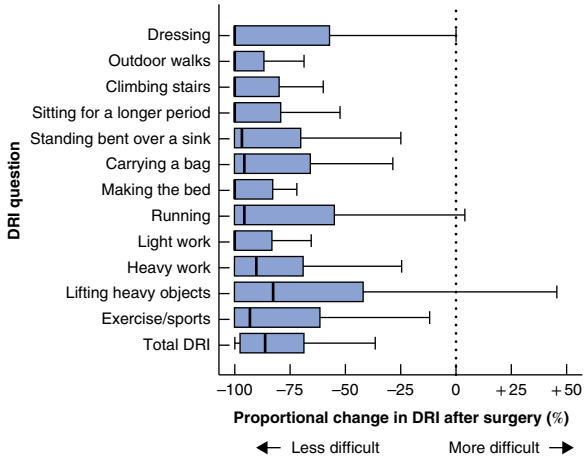
Nineteen of the 60 women (32 per cent) underwent surgical method A, 31 (52 per cent) had method B, and ten (17 per cent) method C. There was at least one concomitant midline fascial defect in 45 women (75 per cent), of which six (13 per cent) were diagnosed at surgery. The correlation coefficient between the rectus diastasis measured by

Table 2 Abdominal Trunk Function Protocol, Urogenital Distress Inventory and Incontinence Impact Questionnaire results before and 1 year after surgery for diastasis recti abdominis

	Before surgery* (n = 60)	After surgery* (n = 60)	P#
Abdominal Trunk Function Protocol			
Specific DRI (0–100 points)†§			
Dressing	1 (0·0–7·0)	0 (0·0–1·0)	0·006
Outdoor walks	4 (0·0–9·5)	0 (0·0–4·4)	< 0·001
Climbing stairs	3 (0·0–9·4)	0 (0·0–4·4)	< 0·001
Sitting for a longer period	28 (0·0–9·6)	0 (0·0–5·2)	< 0·001
Standing bent over a sink	41 (0·0–10·0)	0 (0·0–5·3)	< 0·001
Carrying a bag	29 (0·0–8·3)	1 (0·0–4·9)	< 0·001
Making the bed	13 (0·0–8·4)	0 (0·0–4·5)	< 0·001
Running	49 (0·0–10·0)	1 (0·0–9·6)	< 0·001
Light work	21 (0·0–10·0)	0 (0·0–5·0)	< 0·001
Heavy work	64 (0·0–10·0)	5 (0·0–9·8)	< 0·001
Lifting heavy objects	63 (0·1–10·0)	9 (0·0–9·8)	< 0·001
Exercise/sports	54 (0·1–10·0)	5 (0·0–9·4)	< 0·001
Total DRI score (0–120 points)‡	386(247)	82(118)	< 0·001**
Physiological tests¶			
Back muscle strength (s)†	75 (0–240)	113 (0–240)	< 0·001
Abdominal muscle strength (s)†	49 (0–240)	66 (15–240)	< 0·001
Core stability, side plank (s)†	40 (0–120)	56 (10–115)	< 0·001
Core muscle strength and stability test (s)†	60 (0–180)	74 (3–180)	0·004
Difficulties with active straight leg raising (1–5 points)†	1 (1–5)	1 (0–2)	< 0·001
Pain during straight leg raising	8 (13)	3 (5)	0·096††
Pelvic tip during straight leg raising	9 (15)	9 (15)	1·000††
Pain during pelvic provocation	12 (20)	3 (5)	0·020††
Urogenital Distress Inventory (UDI-6)†	5 (0–16)	2 (0–13)	0·001
Incontinence Impact Questionnaire (IIQ-7)†	2 (0–18)	0 (0–17)	0·002

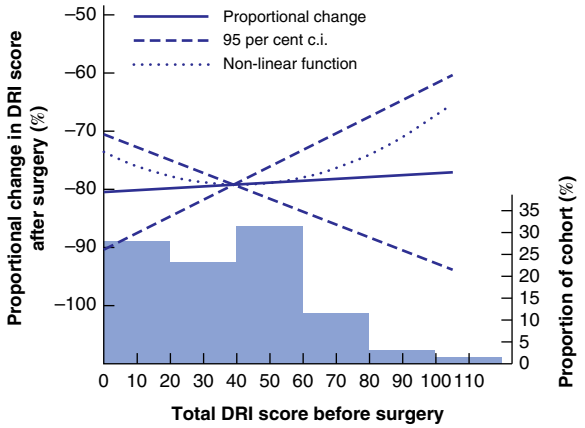
*With percentages in parentheses unless indicated otherwise; values are †median (range) and ‡mean(s.d.). §The Disability Rating Index (DRI) was standardized and recorded on visual analogue scales (measured in millimetres), providing a score with a range of 0–100 for each activity where 0 represented no difficulty at all in performing the specified task and 100 indicated not being able to perform the task at all). ¶ Physiological tests were conducted and monitored by a physiotherapist. #Wilcoxon signed rank test, except **paired t test and ††McNemar test.

Fig. 2 Box plot of the proportional change in the Disability Rating Index score at 1 year *versus* before surgery



Median values and interquartile ranges are denoted by horizontal bars and boxes respectively; error bars have been drawn to span all data points within 1·5 i.q.r. of the nearer quartile. Outliers have been excluded. DRI, Disability Rating Index.

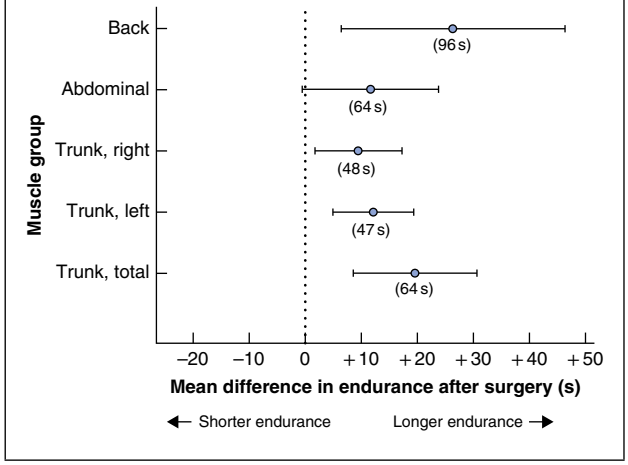
Fig. 3 Proportional change in Disability Rating Index score after surgery



Proportional change in mean total Disability Rating Index (DRI) score after surgery as a function of the preoperative score in a linear regression model. $P_{\text{non-linearity}} = 0{\cdot}740$. The histogram represents the preoperative score distribution.

ultrasonography and the intraoperative finding was 0·71. The corresponding coefficient for CT was 0·55. In general, ultrasound imaging tended to underestimate the mean diastasis by 4 mm ($P = 0{\cdot}007$) and CT overestimated by 3 mm ($P = 0{\cdot}139$). The median hospital stay was 3 (range 1–8) days.

Fig. 4 Mean change in endurance of the various physical tests before and after surgery



Tests were standardized and evaluated by a physiotherapist. Values in parentheses represent the mean score of each variable before surgery.

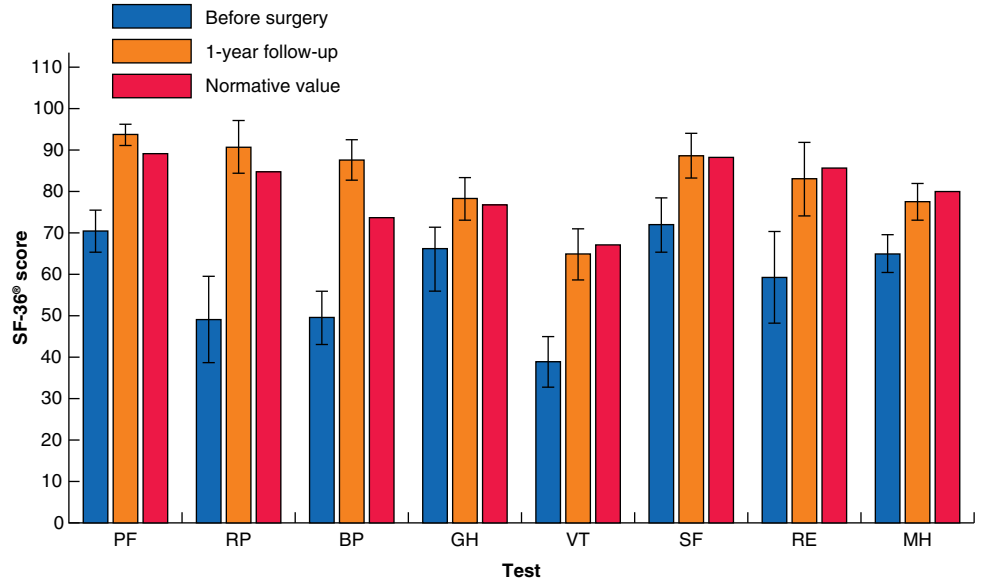
Postoperative complication and recurrence rates

At the 6–8-week follow-up, seven women had a postoperative complication. Four women (3 operated on with method B and 1 with method C) developed bleeding/haematoma that needed reoperation (Clavien–Dindo grade IIIb[33,34]). Two patients (operated on with method B) developed a surgical-site infection requiring antibiotic treatment (Clavien–Dindo grade II). One patient (operated on with method A) presented with spontaneous pneumothorax not requiring intervention 2 weeks after surgery (Clavien–Dindo grade I). Four patients (3 operated on with method B and 1 with method C) developed a seroma not requiring intervention, diagnosed at clinical follow-up 6–8 weeks after surgery (Clavien–Dindo grade I). Finally, four women were not satisfied with the cosmetic result due to umbilical asymmetry, of whom two had reoperation; this was not considered a complication. None of the early complications had led to long-term sequelae at the 1-year follow-up. Complications according to the Clavien–Dindo classification[33,34] were in summary: grade I, five of 60 (8 per cent); grade II, two of 60 (3 per cent); grade IIIb, four of 60 (7 per cent). No recurrences were observed at 6–8 weeks or at 1-year follow-up.

Abdominal Trunk Function Protocol

Table 2 summarizes the ATFP findings before surgery and at 1-year follow-up. Regarding the DRI, 98 per cent of women (55 of 56) reported fewer problems after surgery, and the total scores were, on average, 79·1 (95 per cent

Fig. 5 Bar chart showing Medical Outcome Survey Short Form 36 (SF-36®) scores for patients with symptomatic diastasis recti abdominis



Mean SF-36® scores before and 1 year after surgery are compared with normative values from 3994 women aged 15–64 years in the Swedish SF-36 Health Survey. Error bars indicate 95 per cent confidence intervals. PF, physical function; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health.

c.i. 73·1 to 85·1) per cent lower at follow-up than before surgery. One patient reported a higher score after surgery (total DRI 98 before surgery *versus* 105 after surgery). Median scores and proportional change after surgery for each specific question are displayed in *Fig. 2*. The preoperative score was not associated with proportional change in DRI at follow-up ($P = 0{\cdot}804$) (*Fig. 3*). When evaluated by a physiotherapist, a majority of patients (38 of 50, 76 per cent) had significantly better performance and stamina at follow-up than before surgery. There was no significant change in pain and pelvic tip during straight-leg raising. Although a significant proportion of the women performed better in the postoperative tests, the mean abdominal strength did not appear to have improved at 1-year follow-up (*Fig. 4*).

Quality of life

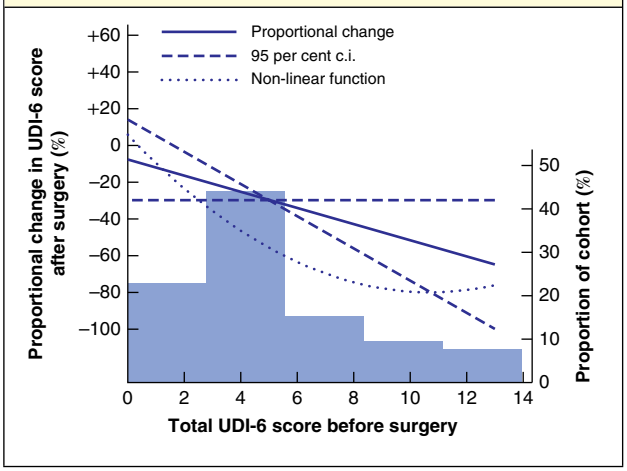
Mean SF-36® subscale scores, comparing results before and after surgery, and with expected ratings in a normative Swedish female population, are shown in *Fig. 5*. Before surgery, the women generally had a lower QoL than the normative Swedish female population in all SF-36® subscales ($P < 0{\cdot}003$). After surgery, their QoL improved significantly, with scores similar to those of the normative Swedish female population in all subscales, and even higher in terms of bodily pain ($P < 0{\cdot}001$).

Urinary incontinence

Table 2 summarizes urinary incontinence before surgery and at 1-year follow-up. A general decrease in incontinence symptoms was observed after surgery, with 28 of the 60 women (47 per cent) reporting a lower score in the UDI-6 after surgery, and eight (13 per cent) reporting a higher score. The mean reduction in score after surgery was 34·4 (95 per cent c.i. 16·4 to 52·3) per cent. There appeared to be an inverse linear relationship between scores before and after surgery, with each point scored before surgery related to a 0·39-point decrease in UDI-6 after surgery. Although no formal evidence of non-linearity was observed ($P = 0{\cdot}062$), no relief of symptoms was observed after surgery when 2 or fewer points were scored before surgery. Otherwise, the proportional decrease in UDI-6 score remained between 40 and 80 per cent (*Fig. 6*).

For the IIQ-7, 22 of the 60 women (37 per cent) had fewer symptoms and five (8 per cent) experienced more symptoms.

Fig. 6 Proportional change in Urogenital Distress Inventory score after surgery



Proportional change in mean total Urogenital Distress Inventory (UDI-6) score after surgery as a function of the preoperative score in a linear regression model. $P_{\text{non-linearity}} = 0{\cdot}062$. The histogram represents the preoperative score distribution.

Discussion

DRA is a potentially debilitating condition in postpartum women that correlates with trunk instability, urinary incontinence and impaired QoL. In this study, a significant improvement in self-reported disability and physical performance, higher QoL and reduced urinary incontinence was observed in the majority of women after surgical treatment of DRA. A novel multimodal protocol to evaluate abdominal trunk function was also introduced.

All women reported a better QoL (SF-36® findings) after surgery, reaching levels similar to those in a normative Swedish female population, regardless of performed surgical method. This indicates that the selection process was successful in distinguishing between physical and cosmetic reasons for surgery. The ATFP focused on functional disability that was resistant to training, thereby selecting patients likely to benefit from surgery. The fact that proportional improvements in DRI and physical tests were similar for all patients, regardless of preoperative scores, implies that surgical reconstruction leads to improvement in function in all postpartum women with DRA. With respect to the postoperative rehabilitation programme, it is not likely that rehabilitation with a lower load than the preoperative training would have had any significant impact on the improvement.

It is suggested that patients with DRA not causing dysfunction (DRA less than 30 mm, midline hernia or cosmetic issues) should first and foremost receive conservative management with weight control, limited hernia repair or purely aesthetic surgery. Evidence in the literature supporting physical training for symptomatic DRA[18–21] is inconclusive. The main purpose of physical training is possibly to restore function, and not necessarily to reduce the diastasis. If physical training proves to be unsuccessful, surgical reconstruction may be the next step in the treatment algorithm.

The improvement in urinary incontinence symptoms observed in this study is in line with previous studies[26], and may indicate a correlation between abdominal trunk instability and pelvic floor dysfunction. Higher preoperative UDI-6 scores resulted in greater improvements in urinary incontinence symptoms after surgery, suggesting that patients with severe symptoms benefit more from surgical reconstruction than those with mild symptoms – an important factor when selecting patients for surgery.

All three surgical techniques used in this study provided similar results regarding outcome and recurrence. None of the women had signs of recurrence 1 year after surgery, and the complication rate was similar to that following other medium-sized surgery such as open ventral hernia repair. There were no differences in QoL outcomes between the three surgical methods, indicating that cosmetic improvement alone was unlikely to be the reason for the improvement in QoL.

The high incidence of perioperative ventral hernia in the study sample could indicate that these hernias contributed to the symptoms; however, the presenting symptoms are not usually associated with a ventral hernia and it is unlikely that the concurrent hernia repair alone could explain the results.

This study has limitations. It lacked a conservatively managed control group, which makes any far-reaching conclusion difficult as some beneficial effects could have been a placebo effect or simply due to the passing of time, although comparison of preoperative and postoperative results allows within-person changes to be measured. The inclusion criteria were restricted to patients with symptomatic persistent DRA, and these results are thus not applicable to all patients with postpartum DRA.

During pregnancy, women are at risk of developing persistent symptomatic DRA that may have a detrimental effect on their physical function and QoL. This study has demonstrated that surgical reconstruction of DRA in postpartum women with symptoms resistant to training results in significant improvements in abdominal trunk function, urinary incontinence and QoL for a majority of patients. Surgical reconstruction of DRA is a valid alternative for patients presenting with symptomatic DRA, where adequate physical training has proven unsuccessful.

Acknowledgements

The authors acknowledge physiotherapist L. Nordstrand for thorough functional examinations, and L. Bai for her illustrations of the surgical procedures employed.

This work was funded by grants from Stockholm County Council and the Swedish Society of Medicine.

Disclosure: The authors declare no conflict of interest.

References

1 Boissonnault JS, Blaschak MJ. Incidence of diastasis recti abdominis during the childbearing year. *Phys Ther* 1988; **68**: 1082–1086.
2 Noble E. *Essential Exercises for the Childbearing Year* (3rd edn). Houghton Mifflin: Boston, 1988.
3 Beer GM, Schuster A, Seifert B, Manestar M, Mihic-Probst D, Weber SA. The normal width of the linea alba in nulliparous women. *Clin Anat* 2009; **22**: 706–711.
4 Rath AM, Attali P, Dumas JL, Goldlust D, Zhang J, Chevrel JP. The abdominal linea alba: an anatomo-radiologic and biomechanical study. *Surg Radiol Anat* 1996; **18**: 281–288.
5 Emanuelsson P, Dahlstrand U, Strömsten U, Gunnarsson U, Strigård K, Stark B. Analysis of the abdominal musculo-aponeurotic anatomy in rectus diastasis: comparison of CT scanning and preoperative clinical assessment with direct measurement intraoperatively. *Hernia* 2014; **18**: 465–471.
6 Akram J, Matzen SH. Rectus abdominis diastasis. *J Plast Surg Hand Surg* 2014; **48**: 163–169.
7 Sperstad JB, Tennfjord MK, Hilde G, Ellström-Engh M, Bø K. Diastasis recti abdominis during pregnancy and 12 months after childbirth: prevalence, risk factors and report of lumbopelvic pain. *Br J Sports Med* 2016; **50**: 1092–1096.
8 Fernandes da Mota PG, Pascoal AG, Carita AI, Bø K. Prevalence and risk factors of diastasis recti abdominis from late pregnancy to 6 months postpartum, and relationship with lumbo-pelvic pain. *Man Ther* 2015; **20**: 200–205.
9 Gitta S, Magyar Z, Tardi P, Füge I, Járomi M, Ács P *et al.* Prevalence, potential risk factors and sequelae of diastasis recti abdominis. *Orv Hetil* 2017; **158**: 454–460.
10 Candido G, Lo T, Janssen PA. Risk factors for diastasis of the recti abdominis. *Journal of the Association of Chartered Physiotherapists in Women's Health* 2005; **97**: 49–54.
11 Hills NF, Graham RB, McLean L. Comparison of trunk muscle function between women with and without diastasis recti abdominis at 1 year postpartum. *Phys Ther* 2018; **98**: 891–901.
12 Parker M, Millar L, Dugan S. Diastasis rectus abdominis and lumbo-pelvic pain and dysfunction – are they related? *J Womens Health Phys Therap* 2009; **33**: 15–22.
13 Spitznagle TM, Leong FC, Van Dillen LR. Prevalence of diastasis recti abdominis in a urogynecological patient population. *Int Urogynecol J Pelvic Floor Dysfunct* 2007; **18**: 321–328.
14 Doubkova L, Andel R, Palascakova-Springrova I, Kolar P, Kriz J, Kobesova A. Diastasis of rectus abdominis muscles in low back pain patients. *J Back Musculoskelet Rehabil* 2018; **31**: 107–112.
15 Gutke A, Ostgaard HC, Oberg B. Association between muscle function and low back pain in relation to pregnancy. *J Rehabil Med* 2008; **40**: 304–311.
16 Ostgaard HC, Zetherström G, Roos-Hansson E. Back pain in relation to pregnancy: a 6-year follow-up. *Spine (Phila Pa 1976)* 1997; **22**: 2945–2950.
17 To WW, Wong MW. Factors associated with back pain symptoms in pregnancy and the persistence of pain 2 years after pregnancy. *Acta Obstet Gynecol Scand* 2003; **82**: 1086–1091.
18 Benjamin DR, van de Water AT, Peiris CL. Effects of exercise on diastasis of the rectus abdominis muscle in the antenatal and postnatal periods: a systematic review. *Physiotherapy* 2014; **100**: 1–8.
19 Mesquita LA, Machado AV, Andrade AV. Physiotherapy for reduction of diastasis of the recti abdominis muscles in the postpartum period. *Rev Bras Ginecol Obstet* 1999; **21**: 267–272.
20 Mommers EHH, Ponten JEH, Al Omar AK, de Vries Reilingh TS, Bouvy ND, Nienhuijs SW. The general surgeon's perspective of rectus diastasis. A systematic review of treatment options. *Surg Endosc* 2017; **31**: 4934–4949.
21 Lee D, Hodges PW. Behavior of the linea alba during a curl-up task in diastasis rectus abdominis: an observational study. *J Orthop Sports Phys Ther* 2016; **46**: 580–589.
22 Toranto IR. The relief of low back pain with the WARP abdominoplasty: a preliminary report. *Plast Reconstr Surg* 1990; **85**: 545–555.
23 Oneal RM, Mulka JP, Shapiro P, Hing D, Cavaliere C. Wide abdominal rectus plication abdominoplasty for the treatment of chronic intractable low back pain. *Plast Reconstr Surg* 2011; **127**: 225–231.
24 Emanuelsson P, Gunnarsson U, Dahlstrand U, Strigård K, Stark B. Operative correction of abdominal rectus diastasis (ARD) reduces pain and improves abdominal wall muscle strength: a randomized, prospective trial comparing retromuscular mesh repair to double-row, self-retaining sutures. *Surgery* 2016; **160**: 1367–1375.
25 Nahas FX, Augusto SM, Ghelfond C. Should diastasis recti be corrected? *Aesthetic Plast Surg* 1997; **21**: 285–289.
26 Taylor DA, Merten SL, Sandercoe GD, Gahankari D, Ingram SB, Moncrieff NJ *et al.* Abdominoplasty improves low back pain and urinary incontinence. *Plast Reconstr Surg* 2018; **141**: 637–645.
27 Salén BA, Spangfort EV, Nygren AL, Nordemar R. The Disability Rating Index: an instrument for the assessment of disability in clinical settings. *J Clin Epidemiol* 1994; **47**: 1423–1435.
28 Biering-Sørensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. *Spine (Phila Pa 1976)* 1984; **9**: 106–119.

29 Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: a validity study of a McKenzie evaluation and sacroiliac provocation tests. *Aust J Physiother* 2003; **49**: 89–97.
30 Sullivan M, Karlsson J, Ware JE Jr. The Swedish SF-36 Health Survey – I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. *Soc Sci Med* 1995; **41**: 1349–1358.
31 Franzén K, Johansson JE, Karlsson J, Nilsson K. Validation of the Swedish version of the incontinence impact questionnaire and the urogenital distress inventory. *Acta Obstet Gynecol Scand* 2013; **92**: 555–561.
32 Emanuelsson P. *Alternatives in the Treatment of Abdominal Rectus Muscle Diastasis: an Evaluation;* 2014. https://openarchive.ki.se/xmlui/handle/10616/42245 [accessed 15 January 2019].
33 Clavien PA, Barkun J, de Oliveira ML, Vauthey JN, Dindo D, Schulick RD *et al.* The Clavien–Dindo classification of surgical complications: five-year experience. *Ann Surg* 2009; **250**: 187–196.
34 Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004; **240**: 205–213.

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article.

BMJ Open

Perceived barriers to multiprofessional team briefings in operating theatres: a qualitative study

Laura Fruhen ,[1] Joseph Alexandre Carpini ,[2] Sharon K Parker,[3] Yee Leung,[4] Adrian F S Flemming[5]

To cite: Fruhen L, Carpini JA, Parker SK, *et al*. Perceived barriers to multiprofessional team briefings in operating theatres: a qualitative study. *BMJ Open* 2020;**10**:e032351. doi:10.1136/bmjopen-2019-032351

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-032351).

Received 17 June 2019
Revised 22 November 2019
Accepted 30 December 2019

For numbered affiliations see end of article.

Correspondence to
Dr Laura Fruhen;
laura.fruhen@uwa.edu.au

ABSTRACT

Objectives This study investigates perceived barriers towards the implementation of multiprofessional team briefings (MPTB) in operating theatres, as well as ways to overcome these perceived barriers. Previous research shows that MPTB can enhance teamwork and communication, but are underused in operating theatres. By adopting a multilevel systems perspective, this study examines perceived barriers and solutions for MPTB implementation.

Design Participants completed open-ended survey questions. Responses were coded via qualitative content analysis. The analysis focused on themes in the responses and the systems level at which each barrier and solution operates.

Setting Four tertiary hospitals in Australia.

Participants 103 operating theatre staff, including nurses, surgeons, anaesthetists, technicians and administrators.

Results Participants identified barriers and solutions at the organisational (15.81% of barriers; 74.10% of solutions), work group (61.39% of barriers; 25.09% of solutions) and individual level (22.33% of barriers; 0% of solutions). Of all the perceived barriers to MPTB occurrence, a key one is getting everyone into the room at the same time . Matching of perceived barriers and solutions shows that higher systems-level solutions can address lower level barriers, thereby showing the relevance of implementing such wider reaching solutions to MPTB occurrence (including work practices at occupational level and above) as well as addressing more local issues.

Conclusions Successful MPTB implementation requires changes at various systems levels. Practitioners can strategically prepare and plan for systems-based strategies to overcome barriers to MPTB implementation. Future research can build on this study's findings by directly examining higher systems-level barriers and solutions via detailed case analyses.

Strengths and limitations of this study

- This study investigates barriers and solutions to the implementation of multiprofessional team briefings (MPTB), as perceived by operating theatre staff (n=103).
- Open-ended questions allowed participants to freely bring up topics that were salient to them.
- Inductive analysis of participants' responses adopted a multilevel systems model to identify barriers and solutions to MPTB implementation.
- The sample was of sufficient size and represented key stakeholders in operating theatres.
- This study is only descriptive and there may be differences between perceived and actual barriers to MPTB implementation.

BACKGROUND

The purpose of this study is to identify the perceived barriers and solutions towards the successful implementation of multiprofessional team briefings (MPTB) in operating theatres (OTs). Preoperative communication between staff is not well studied, yet is reported as underused and lacking a standard method or procedure.[1] MPTB serve as a potential standardised complement to surgical checklists that can enhance preoperative communication and hence theatre performance and safety.[2] In OTs across the globe, surgical checklists have been widely adopted. The introduction of checklists has had clear benefits to patient outcomes in terms of improved detection of safety hazards, reduced postoperative mortality and reduced complication rates.[3 4] Nonetheless, there is still room for further improvement in how surgical teams work together. Globally, 16.8% of patients undergoing elective surgery develop one or more postoperative complication and 0.5% die.[5] The fact that preventable complications still occur in OTs despite the implementation of surgical checklists suggests that surgical checklists alone cannot fully address the dynamic issues that contribute to negative patient and surgical team outcomes.[1 6]

Preoperative communication and MPTB specifically have been identified as a complementary approach to checklists that provides opportunity for team building and addresses sociocultural aspects related to teamwork that checklists do not directly address. Importantly,

MPTB offer the opportunity for teams to develop shared mental models, effective team work behaviours and communication.[2] MPTB are short meetings, conducted at the beginning of a surgical list before the first patient arrives.[7] This timing has been rated favourably by surgical staff, compared with the timing of the checklists immediately prior to the start of each procedure.[1] They include those surgeons, nurses, anaesthetists and technicians involved in an OT list. The purpose of MPTB is to enable theatre staff to share information, to create a team identity where information can be freely exchanged and to plan ahead across the full list. In this way, MPTB are distinct from and complement checklists and may address some of the underlying issues still contributing to preventable surgical complications and never events.[8] MPTB at the start of operating lists vary with regard to their structure and content. Leong *et al*[7] followed three steps (introduction round, tasks of the team members and expected technical or logistical issues that require extra attention). Bleakley *et al*[9 10] describe a typical MPTB at the start of an OT list as consisting of a technically-oriented discussion of the list led by the surgeon, equipment checks, patient lists and potential problems. In another study, team members first introduced themselves and their roles, develop a plan for the day and discuss critical aspects of each procedure as an MPTB.[11] Bethune *et al*[12] describe MPTB as including feedback from previous lists, consideration of external factors and a discussion of each patient on the list. Specific to our research, staff working in the hospitals involved in this study were encouraged to cover five steps in MPTB: (1) staff introduction including name and team role, (2) overview of the surgical list (eg, half-day/full day, number of patients), (3) relevant details of each case, (4) questions and (5) summary of any changes or issues discussed.[6]

MPTB have been recognised as having positive effects on theatre outcomes. Studies show that they benefit teamwork[7 9–14] safety,[7 9–11 13] as well as efficiency.[7 12] More generally, lack of standard methods for preoperative team communication has been identified,[1] and MPTB are ideally placed to address this gap. Yet, they still remain an infrequent and underused practice in Australian OTs.[2 13] Issues with uptake, or even resistance, are reported for other teamwork interventions in OTs (ie, checklists[15–19]). These include resistance from professionals in OTs,[15–17] leadership and established hierarchies,[17 20 21] lack of education[18] and poor communication.[17 19]

Given their relevance to teamwork and surgical outcomes, it is important to understand what is hindering and helping the uptake of MPTB in OTs. Grol and Grimshaw[22] identified that changes in clinical practice are only partially within the control of medical staff. There is a wider recognition in implementation research that implementation strategies need to be carefully targeted towards barriers or obstacles.[22 23] Accordingly, this study investigates the perceived barriers to the introduction of MPTB and solutions for overcoming these barriers. In recognition of the need for a systems approach towards change implementation and integration to work in OTs[24] and healthcare more widely (see process normalisation theory,[25]) we apply Parker *et al*'s six-level model of work design influences[26] to categorise the perceived barriers and solutions identified by surgical professionals. Parker *et al*[26] identify six multilevel influences that interact with, and shape each other to affect work design in a particular situation. The six levels of influence on how work is organised within any particular situation include: global factors (eg, WHO strategies, global migration patterns that affect staffing), national factors (eg, Gross Domestic Product (GDP) and other economic aspects, government healthcare policies, industrial relations policies), occupational factors (eg, professional norms), organisational factors (eg, organisational design and culture), work group factors (eg, composition, local leadership) and individual factors (eg, education, motivation). To change work design, implementing MPTB in this case, or at least sustain a change in work design, one needs to adopt a full systems perspective to understand how multiple levels of influence shape work design and staff behaviours and how they interact with each other. Doing so recognises the complex nature of implementing quality improvement initiatives and can support individuals and organisations in their efforts to incorporate MPTB into standard practice.

METHODS

OT staff from four tertiary metropolitan hospitals were surveyed on their perceptions of perceived barriers towards conducting MPTB and the solutions for overcoming these perceived barriers. MPTB were encouraged at all sites; however, they were not fully integrated into daily practice. As part of presentations about MPTB at staff meetings, participants were asked to note down challenges they experienced in conducting MPTB in their workplaces, and to propose potential solutions to these challenges. A second wave of data collection was conducted via online surveys specifically addressed at surgeons and OT administrators as previous attempts to engage these groups were unsuccessful. Responses were provided in an open-ended text format and participants could provide as many or as few barriers and solutions as they wish (see online supplementary file provided). Participation was voluntary and anonymous. Participants were informed of the ways in which their responses would be used.

Sample

A sample of n=103 OT staff (out of which 44 were nurses, 13 technicians, 16 anaesthetists; 20 surgeons; 4 administrators; note that the remainder seven of the sample did not indicate their professional group) participated in this study. The data were collected over a 3-year period (wave 1; 2014–2017) and all OT staff were invited to contribute to this work on several occasions. Almost all nurses and anaesthetists attended information and training sessions

where the opportunity to participate was given. Information and training sessions were not attended by surgeons and thus more direct methods of recruitment were necessary (eg, direct contact and surgical meeting attendance; wave 2, 2019). The sample was of sufficient size for saturation to occur in our analysis.[27]

Patient and public involvement

No patients or members of the public were involved in this study.

Data analysis

Responses were coded independently by two raters following an inductive coding framework.[28] Both raters were trained psychologists with a background in industrial and organisational psychology. The analysis was conducted in three steps. In the first step, one rater read all responses and identified emergent themes. In the second step, each response was assigned to one of the themes that had emerged during the first step. This second step involved two raters independently analysing the responses. Inter-rater reliability was assessed using Krippendorff's alpha and indicated that the raters were highly consistent in their coding of responses (α_{Kripp}=0.87, $CI_{95\%LL}$ 0.80, $CI_{95\%UL}$ 0.94[29]). Results were generated by frequency counts per emergent theme, which is an indication of perceived relevance.[30] In the final step, the identified themes were classified into the levels of work design influences by Parker *et al*.[26]

RESULTS

Perceived barriers to the implementation of MPTB in OTs

A total of 214 perceived barriers to implementing MPTB in OTs were identified (Md=2 per participant). These barriers reflected six themes. Table 1 shows the frequency of the themes identified in the perceived barriers and categorises these by systems levels. Notably barriers were predominantly identified at the work group level, one of the lowest systems levels described by Parker *et al* [26] (f=132). Within the work group level, the majority of responses (f=117) focused on staff not being in the OT at the same time at the start of a list and conflicting tasks as a key barrier to MPTB, making this issue the most common barrier to holding briefings. This barrier was sometimes attributed to various reasons such as having different start times, setups or staff being late. Further, communication issues, such as confusion due to information accuracy and specificity (eg, around procedure details, equipment needs), or challenges in interacting constructively with other team members were also reported as a barrier at the work group level (f=15). Next most frequently, perceived barriers were reported at the individual level (f=48).

Table 1 Illustrative quotes and frequencies of each barrier

Theme	Example quotes	Frequency (*f*)	Frequency (% of total)*
Organisational-level barriers		34	15.81
List attributes	Unexpected changes in lists due to emergency cases Having different surgeons throughout the day	20	9.30
Organisational constraints	Organisations audit the start time/briefings not accommodated in schedule Inadequate staffing AM list overruns, affecting PM list start time	14	6.51
Work group-level barriers		132	61.39
Not everyone present/ conflicting tasks	Surgeon or anaesthetists finishing rounds Complex set ups Staff nurses busy locating equipment Late team members Team members are not available before or at 8am	117	54.42
Miscommunication	Confusion over operating surgeon Junior doctors may not know enough of the patient but are the ones representing the consultants at the briefings	15	7.00
Individual-level barriers		48	22.33
Negative attitudes	Briefings [are] done as a formality with steps missing and no space for questions Not interested, refuses to participate Not supportive of the process	38	17.67
Lack of knowledge	Visiting surgeons not knowing procedures Staff present during the briefing may be different to ones involved in specific surgeries Junior staff not being aware of briefings	10	4.65

*Total of 214 responses.

Table 2 Illustrative quotes and frequencies of each solution

Theme	Example quotes	Frequency (*f*)	Frequency (% of total)*
Organisational-level solutions		103	74.10
Scheduling and staffing for MPTB	Ensure staff [are] present before conducting the briefing even if it delays the list Make arrival time earlier Establish and communicate a particular time for all theatres and all staff to conduct briefing	39	28.06
Education of OT staff	Provide statistics to support the benefits [of briefings) Further education for those not participating Signage in each theatre with briefing steps	34	24.46
Organisational policy changes	Making it a hospital policy/audit to ensure all members are present Making it mandatory	13	9.35
Culture change	Encouragement, senior support, management support Need engaged anaesthetic lead	14	10.07
Technology	Call the missing team member	3	2.12
Work group-level solutions		36	25.09
Better communication	Communicate Having team leaders who are good communicators	21	15.12
Enforcing briefings as a priority	Make it a priority Do not bring patient into the OR until briefing is done Preparing beforehand	15	10.79

*Total of 139 responses.

These perceived barriers include lack of knowledge about MPTB (f=10) and negative attitudes towards MPTB (f=38). Such negative attitudes included staff not taking briefings serious and not seeing the benefit of them, as well as active resistance to briefings. Two perceived barriers at the organisational level were reported (f=34), namely surgical list attributes (f=20; eg, lists with only emergency cases, variation in surgical staff) and organisational constraints (f=14; eg, previous list runs over, inadequate staffing). Ccomparisons of frequencies of barriers reported by surgeons and nurses showed identical rankings of the three levels (ie, work group barriers were most frequently identified in the two professional groups, followed by individual and organisational barriers).

Solutions to MPTB implementations

A total of 139 potential solutions for overcoming perceived barriers to implementing MPTB were provided by the participants (Md=1 solution per participant). Within the suggested solutions, a total of seven themes were identified. Table 2 illustrates the frequency and content of each solution. Solutions resided at organisational (f=103) and work group level (f=36). At organisational level, participants suggested changes to staffing and scheduling (f=39) and education around the benefits and procedures for effective MPTB as solutions to MPTB implementation (f=34). A number of solutions at organisational level emerged with relatively low frequency. Participants suggested instituting organisational changes via policies (f=13; eg, making briefings mandatory) and organisational culture change (f=14). Further, use of technology, such as phones, was also mentioned (f=3).

At work group level, two solutions emerged, namely better communication within the OT team to ensure high-quality briefings (f=21) and strategies that allow team members to enforce briefings as a priority within each team (f=15).

DISCUSSION

This qualitative study identifies perceived barriers to the implementation of MPTB in OTs and potential solutions for overcoming these perceived barriers. Investigating what hinders such briefings from occurring is important as MPTB can support effective teamwork and communication in OT teams. Previous research has shown that similar quality improvements have faced resistance in OTs[4] and emphasises a systems approach towards team interventions.[24] To assist operating staff in implementing MPTB as a day-to-day practice, it is important to identify and understand potential barriers that may make the implementation of MPTB more difficult so that solutions can be targeted towards overcoming these specific issues. The present research identifies the barriers and solutions specific to MPTB implementation. In doing so, it can assist practitioners and hospital administrators wanting to implement MPTB into their standard practice. The present study focuses uniquely on the perceived barriers to MPTB implementation as previous research into the implementation of healthcare interventions has

highlighted the necessity to conduct a thorough barrier analysis prior to trying to implement a new practice as these barriers can often result in an intervention failing regardless of the facilitating factors.[23 31] Our findings extend this approach by not only identifying the barriers themselves but also soliciting insights from staff as to the potential solutions. By identifying potential solutions to perceived barriers to implementation, hospitals seeking to implement MPTB will be better equipped to proactively manage potential barriers[32] and can design comprehensive and targeted strategies to address barriers to change.[22]

Barriers to MPTB identified in this study were at organisational, work group and individual levels, with the majority occurring at work group level. A number of these barriers (eg, issues related to attitudes and knowledge) identified in this study are similar to those that have been identified in relation to the implementation of the WHO checklists.[4] The focus at work group level reflects MPTB's status as a team-level work design intervention[6]; however, the emergence of barriers at other levels illustrates the relevance of a systems view on MPTB implementation. Notably, no barriers at the occupational, national or global levels were identified by respondents in this study. Despite this finding, we posit, based on Parker *et al*'s[26] systems model of work design influences and previous research on change implementation in medical settings[3 22] that such barriers exist, but that OT staff were not necessarily aware of them. In contrast to immediate barriers at the individual and work group levels that are likely salient in the day-to-day experience of OT staff, barriers existing at the occupational, national and global levels likely shape MPTB implementation indirectly in more subtle ways that are often difficult to identify. Crucially though, these barriers at occupational, national or global level are wide reaching and the successful adoption of strategies designed to overcome these challenges are likely essential for sustained change.[22 24 26] In the case of MPTB, at occupational level, barriers such as surgical work practices that apply across multiple hospitals (such as the consultancy model of surgeon work), and at national-level issues such as healthcare funding models are likely to be relevant.

Consolidating our results and the above discussion into a process model of systems barriers to MPTB occurrence (based on[26]), figure 1 illustrates the complex interactions of barriers at various systems levels identified in this study and how they can contribute to MPTB occurrence. We identify staff not being in the room and being occupied with conflicting tasks as a core, immediate bottleneck to MPTB occurrence from our findings (based on it being the most frequently identified barrier).

The remaining barriers are likely to affect staff presence for MPTB attendance and to interact with each other. For example, at the individual level, attitudes are likely to be directly linked to staff presence for MPTB but are also likely to adversely affect communication, which in turn can also contribute to staff not being present for MPTB. Further, attitudes are also likely to be shaped by organisational constraints, such as inadequate staffing levels, which can contribute to staff dismissing MPTB, as it can appear like another task they need to engage in.

Participants generated fewer solutions for MPTB implementation than barriers. However, the content of their responses was varied, so that a wide range of solutions could be identified. Solutions for MPTB implementation that participants generated resided at the organisational and work group levels only. Similar to the barriers to MPTB, participants did not report higher system levels solutions at either the occupational or national levels. As has been argued above, barriers and solutions are likely to exit at each level. The criticality of higher level solutions becomes clear when matching barriers and solutions based on their content (see content in grey in figure 1). Content matching illustrates that solutions are likely to address barriers that reside at the same system level or

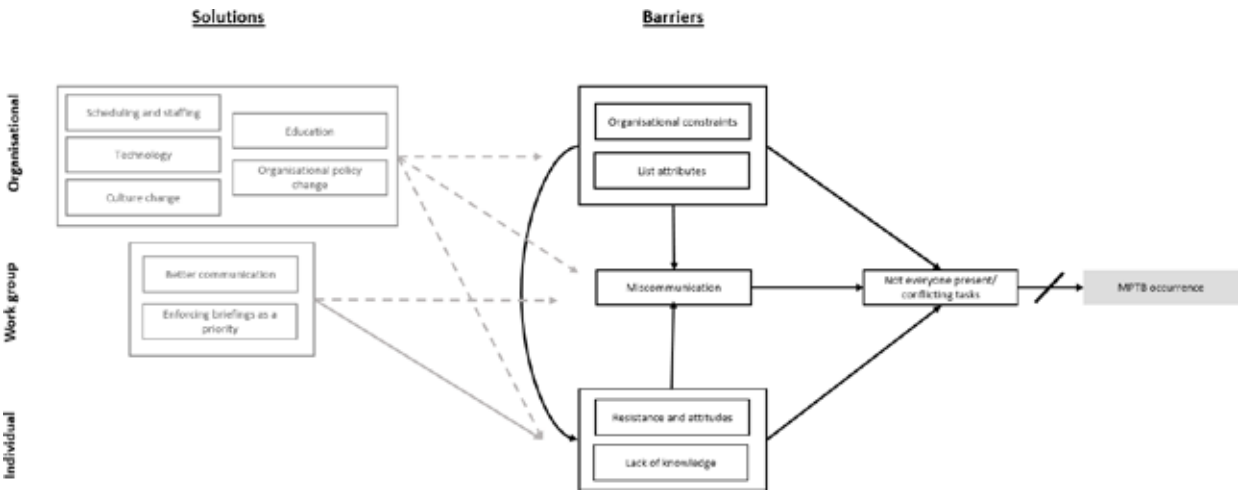


Figure 1 A process model of work design barriers to MPTB occurrence. MPTB, multiprofessional team briefings.

below. However, a solution is unlikely to reach a barrier at a level that exceeds the level of the solution. As is shown in figure 1, solutions at the organisational level are likely to also address barriers at the same level or below, but not above. For example, scheduling and staffing solutions can address organisational constraints (ie, organisational-level barrier), as well as presence of staff at MPTB (work group-level barrier), thereby addressing barriers at organisational and team levels. However, they are unlikely to reach barriers at occupational or national levels. Similarly, education of OT staff can address the individual-level barrier of attitudes by clarifying the benefits of team briefings, and reduce miscommunication at the team level. However, education of OT staff is unlikely to address barriers that reside above the organisational level, such as cost minimisation or staffing levels. While education emerged in our study as an individual-level solution with a focus on OT staff, it needs to be recognised that differently targeted education can also facilitate MPBT implementation at higher system levels. To address barriers at higher levels, concerted efforts can be taken to actively disseminate research findings and educate policy-makers so that best practices can be fully endorsed and adopted by health departments and included in their standards of patient care.

Future research may consider investigating the processes involved in MPTB implementation via case studies and in-depths interviews with OT staff and administrators. Such case studies may identify more detailed information on the barriers and solutions for MPTB implementation identified here. Crucially, our study identifies perceived barriers, however, it is unclear to what extent these overlap with actual barriers to MPTB implementation. Further, longitudinal investigations of barriers and solutions over the course of MPTB implementation may help illuminate the dynamic relationship between barriers and solutions at different levels and provide a process perspective to this type of quality improvement initiatives.

It also needs to be considered that strategies that do not directly target MPTB implementation may also have benefits for their wider implementation success. As many other strategies around communication in healthcare, MPTB are rooted in established practices of crew resource management (CRM) in aviation and other industries.[13 33 34] In addition to the solutions identified in this study, the implementation of MPTB in practice may also benefit from considering other practices from CRM team training targeted at teamwork and communication by growing awareness and appreciation for teamwork efforts like MPTB more generally.

Study strengths and weaknesses

This is the first study to investigate the perceived barriers and solutions to the implementation of MPTB in OTs. The research adopts a multilevel systems approach grounded in theory, which has generated practical guidance and solutions and illustrates the complexity of MPTB implementation. The data were collected via anonymised open-ended survey questions. Using surveys, rather than interviews, allowed for inclusions of a larger group of participants, so that the results are more representative. Responses provided a high level summary that captured the issues well, however, did not generate more in depth reflection as to why participants perceive specific barriers and potential solutions. Further, Grol and Grimshaw[22] describe barriers and facilitators of change in clinical work contexts. However, our study only focused on barriers and examined the strategies to overcome them (ie, solutions), so that other aspects that may need to be considered for successful implementation were not captured. In particular, our study, while addressing one of the key issues associated with implementation processes did not consider facilitators, as issues that may suport the implementation of MPTB in OTs.[22] Such facilitators of change implementations in clinical contexts may include incentives, feedback or perceived social norms.[22] Finally, our study captured the frequency with which barriers and solutions were reported. It needs to be noted that, while frequency in content analysis has been described as a marker of relevance,[30] it may be affected by awareness, or other factors that may lead participants to refer to one issue over another.

Conclusion and implications

Considering the barriers and solutions to MPTB implementation, this paper illustrates that a work design change needs to be built on an understanding of how multiple systems levels shape work designs and behaviours in OTs. Barriers and solutions to MPTB' implementation were predominantly reported by OT staff at work group level, whereas solutions were most likely to reside at organisational level. Notably, our participants did not identify higher level barriers and solutions at occupational or national levels. Yet, our matching of barriers and solutions to MPTB implementation illustrates the possible limitations of lower level solutions in overcoming wider systemic changes in the underlying processes that are necessary to sustain the implementation of MPTB. Our findings reinforce the importance of systems-based change in generating adequate ways of addressing common barriers to MPTB implementation.

Author affiliations
[1]School of Psychological Science, The University of Western Australia, Crawley, Western Australia, Australia
[2]Business School, The University of Western Australia, Crawley, Western Australia, Australia
[3]Future of Work Institute, Curtin University, Perth, Western Australia, Australia
[4]Division of Obstetrics and Gynaecology, Faculty of Health and Medical Sciences, The University of Western Australia, Crawley, Western Australia, Australia
[5]Faculty of Health and Medical Sciences, Surgery, The University of Western Australia, Crawley, Western Australia, Australia

Acknowledgements We thank Fiona Stanley Hospital and Fremantle Hospital, as well as two other hospitals in WA for collaborating with us. We in particular want to express our gratitude to the medical, nursing, administrative and technical staff in the operating theatres who volunteered their valuable time and insights.

Contributors SKP was the principal investigator for this study. LF and JAC distributed the surveys. LF analysed the responses (together with a research assistant) on which the paper is based and led on writing this manuscript. LF, SKP, JAC, AFSF, and YL contributed to the design of the study and editing of the final manuscript and the refining of its intellectual content.

Funding This work was supported by the Government of Western Australia Department of Health grant number (F-AA-33992).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This research was granted ethical approval by the Human Research Ethics Office at the University of Western Australia (approval number/ID: RA/4/1/6933) and was performed in accordance with ethical standards.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. The datasets analysed for the current study are not publicly available, as participants were not asked to consent to the sharing of the data at the time of data collection.

ORCID iDs
Laura Fruhen http://orcid.org/0000-0002-9599-3468
Joseph Alexandre Carpini http://orcid.org/0000-0002-4694-2672

REFERENCES

1 Cruz SA, Idowu O, Ho A, *et al*. Differing perceptions of preoperative communication among surgical team members. *Am J Surg* 2019;217:1–6.
2 Fruhen LS, Carpini J, Parker SK. Let's Talk Teamwork: Multi-professional team briefings in WA operating theatres. *Surgical News, The Royal Australian College of Surgeons* 2017;18:16–18.
3 Abbott TEF, Ahmad T, Phull MK, *et al*. The surgical safety checklist and patient outcomes after surgery: a prospective observational cohort study, systematic review and meta-analysis. *Br J Anaesth* 2018;120:146–55.
4 Treadwell JR, Lucas S, Tsou AY. Surgical checklists: a systematic review of impacts and implementation. *BMJ Qual Saf* 2014;23:299–318.
5 International Surgical Outcomes Study group. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. *Br J Anaesth* 2016;117:601–9.
6 Carpini JA, Flemming AFS, Parker SK. Multidisciplinary team briefings: a way forward. *Day Surgery Australia* 2015;14:12–14.
7 Leong KBMSL, Hanskamp-Sebregts M, van der Wal RA, *et al*. Effects of perioperative briefing and Debriefing on patient safety: a prospective intervention study. *BMJ Open* 2017;7:e018367.
8 NHS Improvement. *Provisional publication of never events reported as occurring between 1 April and 31 October*. London: NHS Improvement, 2018.
9 Bleakley A, Allard J, Hobbs A. Towards culture change in the operating theatre: embedding a complex educational intervention to improve teamwork climate. *Med Teach* 2012;34:e635–40.
10 Bleakley A, Boyden J, Hobbs A, *et al*. Improving teamwork climate in operating theatres:The shift from multiprofessionalismto interprofessionalism. *J Interprof Care* 2006;20:461–70.
11 Khoshbin A, Lingard L, Wright JG. Evaluation of preoperative and perioperative operating room briefings at the hospital for sick children. *Can J Surg* 2009;52:309–15.
12 Bethune R, Sasirekha G, Sahu A, *et al*. Use of briefings and debriefings as a tool in improving team work, efficiency, and communication in the operating theatre. *Postgrad Med J* 2011;87:331–4.
13 Allard J, Bleakley A, Hobbs A, *et al*. "Who's on the team today?" The status of briefing amongst operating theatre practitioners in one UK hospital. *J Interprof Care* 2007;21:189–206.
14 Allard J, Bleakley A, Hobbs A, *et al*. Pre-surgery briefings and safety climate in the operating theatre. *BMJ Qual Saf* 2011;20:711–7.
15 Sparkes D, Rylah B. The world Health organization surgical safety checklist. *Br J Hosp Med* 2010;71:276–80.
16 Sewell M, Adebibe M, Jayakumar P, *et al*. Use of the who surgical safety checklist in trauma and orthopaedic patients. *Int Orthop* 2011;35:897–901.
17 Conley DM, Singer SJ, Edmondson L, *et al*. Effective surgical safety checklist implementation. *J Am Coll Surg* 2011;212:873–9.
18 Levy SM, Senter CE, Hawkins RB, *et al*. Implementing a surgical checklist: more than checking a box. *Surgery* 2012;152:331–6.
19 Fourcade A, Blache J-L, Grenier C, *et al*. Barriers to staff adoption of a surgical safety checklist. *BMJ Qual Saf* 2012;21:191–7.
20 Helmiö P, Takala A, Aaltonen L-M, *et al*. First year with who surgical safety checklist in 7148 otorhinolaryngological operations: use and user attitudes. *Clin Otolaryngol* 2012;37:305–8.
21 Yuan CT, Walsh D, Tomarken JL, *et al*. Incorporating the world Health organization surgical safety checklist into practice at two hospitals in Liberia. *Jt Comm J Qual Patient Saf* 2012;38:254–60.
22 Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *The Lancet* 2003;362:1225–30.
23 Bauer MS, Damschroder L, Hagedorn H, *et al*. An introduction to implementation science for the non-specialist. *BMC Psychology* 2015;3:32.
24 Catchpole K, Russ S. The problem with checklists. *BMJ Qual Saf* 2015;24:545–9.
25 May C. Towards a general theory of implementation. *Implement Sci* 2013;8:18.
26 Parker SK, Van den Broeck A, Holman D. Work design influences: a synthesis of multilevel factors that affect the design of jobs. *Acad Manag Ann* 2017;11:267–308.
27 O'Reilly M, Parker N. 'Unsatisfactory Saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research* 2013;13:190–7.
28 Gioia DA, Corley KG, Hamilton AL. Seeking Qualitative Rigor in Inductive Research:Notes on the Gioia Methodology. *Organizational Research Methods* 2013;16:15–31.
29 Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Meas* 2007;1:77–89.
30 Namey E, Guest G, Thairu L. Data reduction techniques for large qualitative data sets. In: GGaK M, ed. *Handbook for team-based qualitative research*. Plymouth, UK: AltaMira Press, 2008: 137–59.
31 Shojania KG, Grimshaw JM. Evidence-Based quality improvement: the state of the science. *Health Aff* 2005;24:138–50.
32 Bosch M, van der Weijden T, Wensing M, *et al*. Tailoring quality improvement interventions to identified barriers: a multiple case analysis. *J Eval Clin Pract* 2007;13:161–8.
33 O'Connor P, Flin R. Crew resource management training for offshore oil production teams. *Saf Sci* 2003;41:591–609.
34 Salas E, Burke CS, Bowers CA, *et al*. Team training in the skies: does Crew resource management (CRM) training work? *Hum Factors* 2001;43:641–74.

BMJ Open

Postoperative intravenous parecoxib sodium followed by oral celecoxib post total knee arthroplasty in osteoarthritis patients (PIPFORCE): a multicentre, double-blind, randomised, placebo-controlled trial

Qianyu Zhuang,[1,2] Liyuan Tao,[3] Jin Lin,[2] Jin Jin,[2] Wenwei Qian,[2] Yanyan Bian,[2] Yulong Li,[2] Yulei Dong,[2] Huiming Peng,[2] Ye Li,[2] Yu Fan,[2] Wei Wang,[2] Bin Feng,[2] Na Gao,[2] Tiezheng Sun,[2] Jianhao Lin,[2] Miaofeng Zhang,[4] Shigui Yan,[4] Bin Shen,[5] Fuxing Pei,[5] Xisheng Weng [2]

To cite: Zhuang Q, Tao L, Lin J, *et al*. Postoperative intravenous parecoxib sodium followed by oral celecoxib post total knee arthroplasty in osteoarthritis patients (PIPFORCE): a multicentre, double-blind, randomised, placebo-controlled trial. *BMJ Open* 2020;**10**:e030501. doi:10.1136/bmjopen-2019-030501

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-030501).

Received 17 March 2019
Revised 26 November 2019
Accepted 09 December 2019

For numbered affiliations see end of article.

Correspondence to
Dr Xisheng Weng;
wengxishengpumch@126.com

ABSTRACT

Objectives To evaluate the morphine-sparing effects of the sequential treatment versus placebo in subjects undergoing total knee arthroplasty (TKA), the effects on pain relief, inflammation control and functional rehabilitation after TKA and safety.

Design Double-blind, pragmatic, randomised, placebo-controlled trial.

Setting Four tertiary hospitals in China.

Participants 246 consecutive patients who underwent elective unilateral TKA because of osteoarthritis (OA).

Interventions Patients were randomised 1:1 to the parecoxib/celecoxib group or the control group. The patients in the parecoxib/celecoxib group were supplied sequential treatment with intravenous parecoxib 40 mg (every 12 hours) for the first 3 days after surgery, followed by oral celecoxib 200 mg (every 12 hours) for up to 6 weeks. The patients in the control group were supplied with the corresponding placebo under the same instructions.

Primary and secondary outcome measures The primary endpoint was the cumulative opioid consumption at 2 weeks post operation (intention-to-treat analysis). Secondary endpoints included the Knee Society Score, patient-reported outcomes and the cumulative opioid consumption.

Results The cumulative opioid consumption at 2 weeks was significantly smaller in the parecoxib/celecoxib group than in the control group (median difference, 57.31 (95% CI 34.66 to 110.33)). The parecoxib/celecoxib group achieving superior Knee Society Scores and EQ-5D scores and greater Visual Analogue Scale score reduction during 6 weeks. Interleukin 6, erythrocyte sedation rate and C-reactive protein levels were reduced at 72 hours, 2 weeks and 4 weeks and prostaglandin E2 levels were reduced at 48 hours and 72 hours in the parecoxib/celecoxib group compared with the placebo group. The occurrence of adverse events (AEs) was significantly lower in the parecoxib/celecoxib group.

Conclusions The sequential intravenous parecoxib followed by oral celecoxib regimen reduces morphine consumption, achieves better pain control and functional recovery and leads to less AEs than placebo after TKA for OA.

Trial registration number ClinicalTrials.gov (ID: NCT02198924).

Strengths and limitations of this study

- This is the first study to investigate the efficacy and safety of the sequential analgesia regimen of intravenous parecoxib followed by oral celecoxib after total knee arthroplasty surgery.
- The study employed a prospective, randomised, multicentre design.
- This study explored the benefits of prolonged sequential treatment of parecoxib and celecoxib in medium-term function recovery.
- Potential limitations include the need for further validation studies from other institutions outside China, lack of investigation of the long-term (eg, >3 months) effects of the sequential treatment and compromise of the test accuracy of synovial fluid cytokines.

INTRODUCTION

Osteoarthritis (OA) is a chronic degenerative joint disorder which frequently occurs in the elderly.[1] Total knee arthroplasty (TKA), an effective treatment for end-stage knee OA,[2] has been regarded as the most painful orthopaedic surgery due to the weight-bearing characteristics of the knee joint and the high demand of functional exercise post operation.[3] Inadequate pain control is correlated

with prolonged postoperative bed time, increased incidence of pulmonary infection, deep venous thrombosis, pulmonary embolism and poor functional recovery in some patients after TKA.[4 5]

Multimodal analgesia is currently recommended for postoperative pain control after TKA.[6] As opioid tolerance and related side effects are becoming an increasingly significant problem, and even causing public health emergency, great challenges are faced by pain management post TKA.[7 8] Therefore, the value of non-steroidalanti-inflammatory drug (NSAID), especially selective cyclo-oxygenase-2 (COX-2) inhibitors, as an important alternative has become increasingly prominent.[9 10]

In many Chinese institutions, 40 mg parecoxib is routinely administered intravenously two times per day for the first 3 days after surgery, followed by 200 mg celecoxib administered orally two times per day for 2 weeks or longer. Although this sequential therapeutic strategy has been adopted by most Chinese orthopaedic surgeons for its clinical convenience and satisfactory results during clinical observation, high quality evidence is still lacking to support its use and popularisation.

The PIPFORCE study aimed to investigate the sequential analgesic regimen with intravenous parecoxib followed by oral celecoxib for postsurgical analgesia in OA patients undergoing TKA. The primary objective was to evaluate the morphine-sparing effects of the sequential treatment with parecoxib and celecoxib versus placebo in subjects undergoing TKA. Secondary objectives included comparing the sequential treatment versus placebo for their effects on pain relief, inflammation control and functional rehabilitation after TKA and determining the safety profiles of study and control regimens.

MATERIALS AND METHODS

Study design

This was an investigator-initiated, multicentre, double-blind, randomised, placebo-controlled trial. Details of the trial design have been previously published.[11] The project was registered in the ClinicalTrails.gov site. In brief, 246 consecutive patients who underwent elective unilateral TKA because of OA were screened and enrolled in four tertiary care hospitals in China (Peking Union Medical College Hospital as the coordinating centre, West China Hospital of Sichuan University, People's Hospital of Peking University and Second Affiliated Hospital of Zhejiang University College of Medicine) from 1 December, 2014, to 22 September, 2016. The ethical committees of all participating hospitals approved the study before patient recruitment.

Study participants

Inclusion and exclusion criteria were strictly implemented as stated previously,[11] and all patients signed the informed consent form at screening, before any study-specific procedures were conducted. The present study was performed in agreement with the Consolidated Standards of Reporting Trials statement.

Inclusion criteria

Subject eligibility was reviewed and documented by an appropriately qualified member of the investigator's study team before subject inclusion in the study. In addition, subjects must meet all the following inclusion criteria to be eligible for enrolment:

1. Planned elective unilateral total knee arthroplasty because of OA, to be performed under a standardised regimen of general anaesthesia, as specified in this protocol.
2. Evidence of a personally signed and dated informed consent form indicating that the subject (or a legal representative) has been informed of all pertinent aspects of the study.
3. Age above 18 years (male or female).
4. Male and female subjects of childbearing potential agreeing to use an effective method of contraception throughout the study and for 42 days after the last dose of the assigned treatment. A subject is of childbearing potential if, in the opinion of the investigators, he/she is biologically capable of having children and sexually active.
5. Total duration of the surgical procedure of 4 hours or less.
6. American Society of Anesthesiologists (ASA) grade 1 to 3 cases.
7. Willingness and ability to comply with scheduled visits, treatment plan, laboratory tests, standardised rehabilitation scheme and other study procedures.
8. Satisfactory health as determined by the investigators on the basis of medical history and physical exam.
9. Sufficient psychomotor dexterity and cognitive capacity to use intravenous patient-controlled analgesia.
10. Subjects residing close to the hospital may be considered in priority for convenient and sufficient follow-up.

Exclusion criteria

Subjects will be excluded with any of the conditions listed below:

1. Requirement of a revision to a previous knee arthroplasty and/or planned bilateral knee arthroplasties.
2. Requirement of an emergency knee arthroplasty.
3. Addiction to any NSAIDs and opioids.
4. Known hypersensitivity to COX-2 specific inhibitors, sulfonamides, lactose, NSAIDs, opioids or acetaminophen/paracetamol; a history of asthma, urticaria or allergic type reactions after taking aspirin or other NSAIDs.
5. A history of arthritis (ie, rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis), chronic pain (eg, fibromyalgia), metastasis or Paget's disease.
6. Administration of any investigational medication within 30 days prior to the first dose of study medication or plan to receive any investigational drug

other than those described in the protocol during the study.
7. Any known laboratory abnormality, which in the opinion of the investigators, would contraindicate study participation, including alanine aminotransferase (ALT), aspartate aminotransferase (AST), blood urea nitrogen or creatinine ≧1.5 times the upper limits of respective normal reference ranges.
8. Active malignancy of any type, or a history of malignancy (cases with a history of basal cell carcinoma that has been successfully treated can be entered into the study. Those with a history of other malignancies that have been surgically removed, showing no evidence of recurrence for at least 5 years before study enrolment, were also entered into the study).
9. Inflammatory bowel disease (eg, Crohn's disease or ulcerative colitis), chronic or acute renal or hepatic disorder, a significant coagulation defect or any condition which could preclude the use of NSAIDs or COX-2 specific inhibitors.
10. Active or suspected oesophageal, gastric, pyloric channel or duodenal ulceration.
11. Treatment with warfarin or other anticoagulants in the 30 days preceding the first dose of study medication (cardioprotective aspirin, ≦ or 325 mg/day permitted, when the dose has been stable for at least a month prior to entering the study; anticoagulation is permitted when related to the surgery, with such medicines as low molecular weight heparin, including Lovenox and Fragmin.
12. Anticipated or actual requirement of treatment with lithium.
13. ASA grade 4 to 5 cases.
14. A history of a psychiatric disorder requiring new or changing treatment (a subject with a stable psychiatric disorder on therapy may enter the study if no therapeutic changes have been required for the 4 weeks prior to study entry and none are anticipated for the 2-week duration of this study).
15. A history of uncontrolled chronic disease, or a concurrent clinically significant illness or medical condition, which in the investigators' opinion, would contraindicate study participation or confound data interpretation, including but not limited to: uncontrolled hypertension, uncontrolled ischaemic heart disease, uncontrolled cardiac insufficiency, a history of coronary artery bypass graft surgery, a history of heart valve surgery or coronary stent implantation, a history of peripheral vascular disease or cerebrovascular disease, moderate or severe hepatic impairment, fluid retention, heart failure and abdominal pain of unknown aetiology (or study medication could mask symptoms).
16. Any cognitive impairment or other characteristics that would in the investigator's opinion preclude study participation or compliance with protocol mandated procedures.
17. A history of asthma or bronchospasm, requiring treatment with glucocorticoids.
18. A history of alcohol, analgesic or narcotic abuse.
19. Previous randomisation into the study.
20. Being a staff member of an investigational site or a relative to a site staff member.
21. Participation in other studies within 3 months before the beginning of the current trial.
22. Another severe acute or chronic medical or psychiatric condition, or laboratory abnormality that may increase the risk associated with study participation or investigational product administration, may interfere with data interpretation based on investigators' judgement or would render the subject inappropriate for study enrolment.
23. Pregnancy or breastfeeding in females, or males and females of childbearing potential not using effective contraceptives or agreeing to continue effective contraception from screening through 42 days after the last dose of investigational product.

Procedures

The study consisted of three phases: an initial screening phase completed within 30 days prior to randomisation, a 6-week double-blind treatment phase and a 6-week follow-up phase.

In the first phase, the investigators initiated the required screening procedures after obtaining written informed consent. All eligible patients after selection by inclusive/exclusive criteria were assigned in the order of enrolment to their allocated treatment groups according to a computer-generated randomisation sequence.

In the second phase, after screening completion, participants who remained eligible entered a 6-week double-blind randomised treatment period. All participants underwent standard TKA on unilateral side under general anaesthesia. The surgical techniques and anaesthetic regimen used in the four centres were the same, and have been described clearly in a previous report.[11] Patients in the study group were supplied sequential treatment with parecoxib at 40 mg intravenously two times per day (every 12 hours) for the first 3 days post-surgery followed by celecoxib at 200 mg orally two times per day (every 12 hours) for up to 6 weeks, while control patients were administered the corresponding placebo under the same instructions. Patient-controlled intravenous analgesia with morphine was administrated to all participants starting immediately post-anaesthesia and ending at 24 hours after operation. As long as oral intake is feasible, both groups may receive centrally acting analgesic tramadol hydrochloride in the oral form for rescue analgesia in case of Visual Analogue Scale (VAS) score ≥3. With sufficient pain management, patients were instructed to perform functional exercise according to the standardised post-TKA exercise plan.

In the third phase, a telephone safety follow-up at 12 weeks post-surgery was conducted to record any adverse events that may have occurred. Altogether, there were 10 visits for each participant. Screening was performed at visit 1, and the day of TKA operation was considered day 0. There was a visit 1 day before the operation (visit

2), when patient eligibility was evaluated again, and the visit right after the operation was visit 3. Those on days 1, 2 and 3 post-surgery were regarded as visits 4, 5 and 6, respectively; then there were visits 7, 8 and 9 at 2, 4 and 6 weeks post-surgery, respectively. The last visit, visit 10, was at 12 weeks post-surgery.

Randomisation and blinding

All participants who met the study inclusion and exclusion criteria were randomly assigned in a 1:1 ratio to the parecoxib/celecoxib and placebo groups, respectively. Allocation or randomisation was study site based.

The electronic data capture system automatically generated participant identification numbers in sequence at baseline, which were subsequently linked to treatment assignments at randomisation.

In this trial, a double-blind and imitation design was used to blind patients, treating physicians, investigators and data assessors. All study medications used in the trial were identical in packaging, labelling, usage schedule, appearance, taste and odour.

Ethical review and informed consent

The benefits and risks of patient participation were explained to each patient, legal representative or witness by the investigators or their designees, and signed written informed consent was obtained before the trial. The trial was conducted in accordance with the Declaration of Helsinki.

Outcomes

The primary endpoint was cumulative opioid consumption until 2 weeks post operation, which was calculated as the sum of cumulative morphine consumption over the first postsurgical 24 hours plus opioid consumption until 2 weeks post operation. The conversion equivalent of tramadol to morphine was estimated as 300 mg of tramadol equalling 20 mg of morphine.[11]

The key secondary endpoint was Knee Society Score (KSS) at 6 weeks post operation. Other secondary endpoints included: (1) Western Ontario and McMasterUniversities Osteoarthritis Index (WOMAC),[12] KSS,[13] VAS score and EQ-5D scores[14] prior to operation and at 2, 4 and 6 weeks post operation; (2) cumulative opioid consumption at 24 hours, 48 hours, 72 hours, 4 weeks and 6 weeks post operation.

Exploratory endpoints included: (1) knee circumference (measured 1 cm proximal to the base of the patella); (2) knee skin temperature; (3) peripheral blood, intraoperative intra-articular fluid and postoperative drainage fluid cytokines, including interleukin (IL)-6, IL-8 and prostaglandinE2 (PGE2); (4) erythrocyte sedation rate (ESR) and C-reactive protein (CRP). Safety endpoints included the nature, incidence, duration and severity of adverse events (AEs). AEs occurring during and after trial medication discontinuation and their relationships with study treatment were assessed as well.

Sample size calculation

The primary hypothesis of this trial was that subjects treated with parecoxib/celecoxib would consume less morphine during postoperative 2 weeks. Based on a previous trial,[15] we determined that a total of 86 participants per group would have a 90% statistical power in detecting 100 mg or more in the mean difference of cumulative opioid consumption on day 14 between the two groups, assuming a common SD of 200, and a two-sided α of 0.05. This would result in a total of 172 participants. Estimating that 30% of participants would drop out, a sample size of 246 participants was considered to be adequate for this study.

Statistical analysis

The statistician who conducted the analysis was blinded to group allocation. Summary statistics were used to describe the participant characteristics of the trial groups at baseline in the intention-to-treat (ITT) analysis set. The missing data of cumulative opioid consumption was imputed by the multiple imputation method. The results of multiple imputation data were used as a type of sensitivity analysis for comparing cumulative opioid consumption between groups.[16]

For primary endpoint comparison, the ITT analysis was performed to evaluate differences between groups, and effectiveanalysis population (EAP) and per-protocol (PP) analyses were also performed for sensitivity assessment. The primary endpoint did not follow the Gaussian distribution, and was presented as median (IQR) and tested by the Mann-Whitney U test. Bonferroni correction was used to reduce the significance level as 0.05/6=0.0083. The means (95% CIs) of between-group differences of medians were calculated by the bootstrap method (1000 replications). The generalised linear mixed models (GLMM) were also performed for the primary endpoint, including group, gender, age, height and weight as fixed covariates, and different medical centres as random covariates.

For comparing the secondary and exploratory endpoints, continuous data were presented as means (SDs) or medians (IQRs) as appropriate. The secondary endpoints were analysed by the linear mixed model (LMM), adjusted for gender, age, height, weight and different medical centres. The correlation type of different measurement time points was assumed as the first order autocorrelation. Exploratory endpoints were compared by the Mann-Whitney U test, and the significance level was submitted to Bonferroni correction. For safety endpoints, categorical data were presented as counts and percentages, and tested by the Pearson's $\chi 2$ test or Fisher's exact test. The 95% CIs of absolute risk differences between groups were calculated by the Newcombe-Wilson Score method.[17] All statistical analyses were conducted with the statistical package SPSS, V.18.0 (SPSS Inc) and R 3.4.0 software. Besides Bonferroni correction, statistical significance was defined as $p < 0.05$ with two-sided testing.

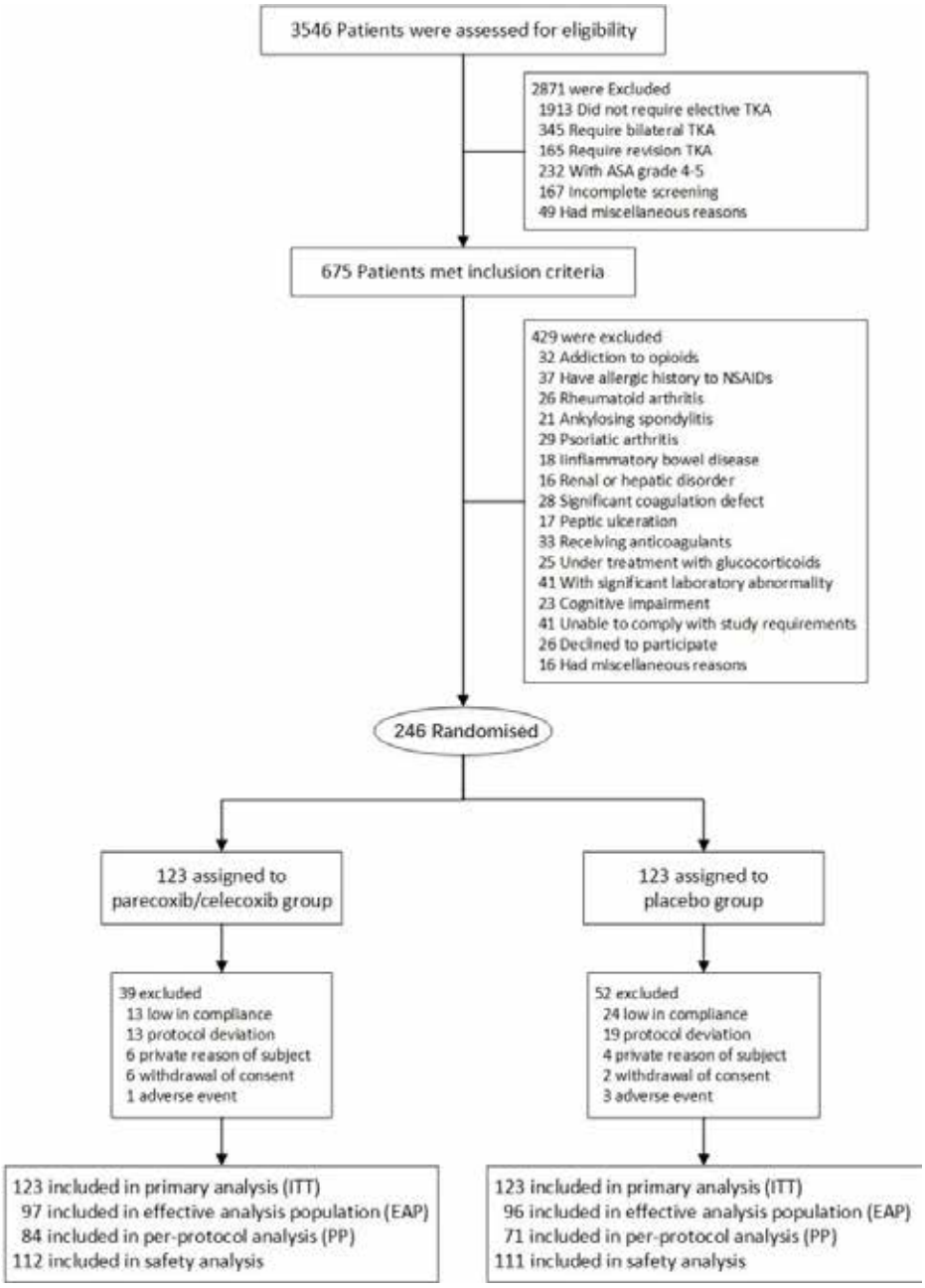


Figure 1 Flowchart of the participants through the study. ASA, American Society of Anesthesiologists; ITT, intention-to-treat; NSAID, non-steroidal anti-inflammatory drug; TKA,total knee arthroplasty.

Quality control and quality assurance

During the study, the investigators or contracted agents performed periodic monitoring visits to ensure Good Clinical Practices. The monitors reviewed all source documents to confirm that the data recorded on case report forms are accurate. The investigators and institutions allowed monitors to directly access source documents for verification.

Each step was strictly performed according to the trial protocol. Each step of quality control of measured outcomes was performed according to the standard operating and quality control procedures.

Patient and public involvement

No patients were involved in conceiving the research question, setting outcome measures or in any other process of the study design. Nor was any patient involved in trial implementation, data collection, data interpretation or writing of the report. There are no plans to

disseminate the results of the study to the subject or the relevant patient communities.

RESULTS

Study patients and follow-up

Patient recruitment began on 1 December, 2014, and the study ended on 6 December, 2016. A total of 3546 participants were screened for eligibility, and 246 patients were ultimately enrolled and randomised (figure 1).

The demographic and baseline characteristics of the randomly assigned patients who received at least one dose of study medication are displayed in table 1 (intention-to-treat set). The baseline characteristics of the two groups were well balanced. There were no statistical differences in age, height and body weight between the placebo and parecoxib/celecoxib groups at baseline.

Primary and secondary outcomes

In ITT analysis, cumulative opioid consumption levels until 2 weeks were significantly reduced in the parecoxib/celecoxib group compared with the placebo group (Z=4.849, p<0.001). The bootstrap method showed that the between-group median difference was 57.31 (95% CI 34.66 to 110.33). The results were similar in EAP and PP analyses (Z=6.619, p<0.001; Z=5.992, p<0.001). Meanwhile, longitudinal analysis by the GLMM showed a significant difference between the two groups (p<0.001) in ITT analysis. Besides, significant opioid consumption reductions throughout postoperative 6 weeks were also observed in the parecoxib/celecoxib group compared with the placebo group (p<0.001; table 2). Sensitivity analysis results from the multiple imputation data set also showed that the placebo group had increased opioid consumption compared with the parecoxib/celecoxib group (online supplementary table 1).

As secondary outcomes, KSS and EQ-5D scores were increased at 2 weeks, 4 weeks and 6 weeks postoperatively in both groups. The LMM showed significant differences between the two groups (p=0.001 and p=0.022, separately), with the parecoxib/celecoxib group achieving superior KSS and EQ-5D scores over the placebo group within 6 postoperative weeks. Similarly, a significant difference between the decreasing tendencies of VAS score was also demonstrated between the two groups (p=0.002). The WOMAC index showed no significant differences between the two groups at the predefined time points (figure 2).

As for the exploratory endpoints, peripheral blood tests revealed that IL-6, ESR and CRP levels were significantly reduced at postoperative 72 hours, 2 weeks and 4 weeks in the parecoxib/celecoxib group compared with the placebo group. PGE2 levels in the intraoperative intra-articular fluid and postoperative drainage fluid, and knee circumference were also significantly reduced at postoperative 48 hours and 72 hours in the treatment group compared with the placebo group. Knee skin

Table 1 Demographic and baseline participant characteristics by group (intention-to-treat analysis)

	Parecoxib/celecoxib (n=123)	Placebo (n=123)
Demographic variables		
Age, mean (SD), y	68.52 (7.26)	67.08 (7.69)
Male, No. (%)	29 (23.58)	20 (16.26)
Female, No. (%)	94 (76.42)	103 (83.74)
Height, mean (SD), cm	159.24 (7.91)	158.07 (6.16)
Weight, mean (SD), kg	65.07 (9.49)	67.65 (11.09)
Clinical variables		
Knee circumference, cm		
Mean (SD)	39.89 (4.30)	40.99 (4.00)
Median (IQR)	40.00 (37.00 to 42.00)	41.00 (38.00 to 43.70)
Knee skin temperature, °C		
Mean (SD)	35.48 (1.46)	35.50 (1.46)
Median (IQR)	36.15 (34.38 to 36.38)	36.17 (35.85 to 36.35)
VAS Score *		
Mean (SD)	5.03 (1.84)	5.35 (1.60)
Median (IQR)	5.00 (4.00 to 6.00)	5.00 (5.00 to 6.00)
WOMAC score*		
Mean (SD)	42.94 (15.56)	44.62 (14.24)
Median (IQR)	43.00 (33.00 to 54.00)	48.00 (34.00 to 55.00)
KSS, mean (SD)*		
Mean (SD)	81.60 (27.29)	75.26 (30.81)
Median (IQR)	81.00 (66.00 to 98.00)	76.63 (56.50 to 92.00)
EQ-5D score, mean (SD)*		
Mean (SD)	0.61 (0.19)	0.58 (0.19)
Median (IQR)	0.66 (0.43 to 0.77)	0.62 (0.41 to 0.73)
Laboratory values		
ESR †		
Mean (SD)	16.16 (12.30)	17.21 (14.26)
Median (IQR)	13.0 (7.0 to 21.0)	13.0 (8.0 to 22.0)
CRP †		
Mean (SD)	3.34 (5.53)	3.14 (3.55)
Median (IQR)	2.21 (1.10 to 3.43)	2.40 (1.23 to 3.81)
PT		
Mean (SD)	11.58 (1.32)	11.56 (1.37)
Median (IQR)	11.40 (10.60 to 12.60)	11.20 (10.60 to 12.50)
APTT		
Mean (SD)	29.97 (5.51)	29.81 (5.64)
Median (IQR)	29.70 (26.10 to 34.70)	28.80 (25.70 to 34.20)
TT‡		
Mean (SD)	17.44 (1.98)	17.87 (2.90)
Median (IQR)	17.30 (16.20 to 18.80)	17.60 (16.20 to 18.80)
FIB		
Mean (SD)	3.04 (0.59)	2.95 (0.67)
Median (IQR)	2.91 (2.58 to 3.45)	2.87 (2.46 to 3.39)

Continued

Table 1 Continued

	Parecoxib/celecoxib (n=123)	Placebo (n=123)

*Data are missing for three participants in the placebo group.
†Data are missing for two participants in the placebo group.
‡Data are missing for two participants in the parecoxib/celecoxib group and one participant in the placebo group.
APTT, activated partial thromboplastin time; CRP, C-reactive protein; ESR, erythrocyte sedation rate; FIB, fibrinogen; KSS, Knee Society Score; PT, prothrombin time; TT, thrombin time;VAS, Visual Analogue Scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

temperature was not significantly different between the two groups (table 3).

Safety

The incidence rate of AEs was significantly lower in the parecoxib/celecoxib group (22.3%) compared with the placebo group (40.5%), and the absolute rate difference between the two groups was −18.22% (95% CI −30.17% to −6.27%; p=0.003). In addition, there were five serious AEs in the placebo group and zero in the parecoxib/celecoxib group (p=0.029). The five serious AEs included one case of joint stiffness, one case of stenocardia, one case of fever and two cases of pain. All of the serious AEs were resolved timely with no sequel after the proper treatment.

No significant differences were detected in AE durations or expected AEs between the two groups (table 4). Other types of adverse events showed no statistically significant differences between the two groups, except that hyperhidrosis, pain, fever, blood glucose and body temperature were increased (table 4).

DISCUSSION

The enhanced recovery after surgery (ERAS) programme[18] has now been recognised and recommended in various elective surgeries.[19 20] The ERAS concept aims to adopt standardised multimodal pathways to improve clinical outcomes, specifically in optimising postsurgical pain control and enabling early rehabilitation.[19 20] Therefore, effective pain management with minimal systemic opioid

Table 2 Cumulative opioid consumption of post operation in two groups

	Parecoxib/celecoxib Median (IQR)	Placebo Median (IQR)	Median difference Median (95% CI) *	P value †
Intention-to-treat	**n=123**	**n=123**		
24 hours	26.13 (24.00 to 32.82)	36.03 (27.63 to 52.00)	10.13 (5.50 to 20.53)	<0.0001‡
48 hours	27.55 (24.01 to 33.60)	45.80 (29.10 to 63.33)	17.74 (6.75 to 28.08)	<0.0001‡
72 hours	28.63 (24.25 to 44.00)	59.57 (29.75 to 88.00)	30.88 (9.24 to 44.27)	<0.0001‡
2 weeks§	44.00 (26.30 to 82.50)	101.80 (42.43 to 199.67)	57.31 (34.66 to 110.33)	<0.0001‡
4 weeks	53.33 (27.17 to 107.17)	166.50 (51.53 to 255.00)	112.02 (43.12 to 150.92)	<0.0001‡
6 weeks	58.00 (30.00 to 116.67)	180.35 (51.53 to 295.00)	120.92 (57.34 to 181.81)	<0.0001‡
Effective analysis population	**n=96**	**n=97**		
24 hours	26.50 (24.02 to 32.75)	38.25 (28.95 to 52.00)	11.95 (5.70 to 21.22)	<0.0001‡
48 hours	27.80 (24.25 to 33.20)	46.54 (29.50 to 64.67)	19.00 (7.92 to 30.32)	<0.0001‡
72 hours	28.75 (24.50 to 42.50)	66.90 (31.56 to 89.33)	37.02 (12.56 to 48.45)	<0.0001‡
2 weeks§	42.98 (26.30 to 80.67)	133.33 (51.53 to 205.00)	88.83 (48.07 to 134.87)	<0.0001‡
4 weeks	51.09 (27.17 to 90.93)	178.42 (64.00 to 265.33)	126.72 (69.16 to 176.09)	<0.0001‡
6 weeks	56.54 (30.00 to 108.75)	190.00 (64.00 to 301.33)	137.71 (106.39 to 197.17)	<0.0001‡
Per-protocol population	**n=84**	**n=71**		
24 hours	26.28 (24.00 to 32.25)	37.50 (28.80 to 52.00)	12.35 (12.05 to 12.65)	<0.0001‡
48 hours	28.13 (24.24 to 34.10)	46.53 (29.87 to 63.33)	18.76 (18.43 to 19.09)	<0.0001‡
72 hours	29.51 (24.50 to 43.25)	65.67 (32.50 to 88.67)	33.08 (32.47 to 33.69)	<0.0001‡
2 weeks§	42.98 (26.28 to 74.50)	133.33 (50.00 to 199.67)	90.95 (89.39 to 92.53)	<0.0001‡
4 weeks	48.93 (26.55 to 90.80)	173.33 (59.17 to 265.33)	124.89 (123.23 to 126.56)	<0.0001‡
6 weeks	56.54 (27.43 to 107.96)	185.33 (59.17 to 307.33)	137.83 (135.73 to 139.92)	<0.0001‡

Data were presented as median (IQR) and tested by the independent Mann-Whitney U test.
*The median difference was placebo group minus parecoxib/celecoxib group, and median (95% CI) was calculated by the bootstrap method (1000 replications).
†The significance level was set as 0.05/6=0.0083 according to the Bonferroni correction.
‡The difference was statistically significant.
§The cumulative opioid consumption until 2 weeks post operation was the primary endpoint.

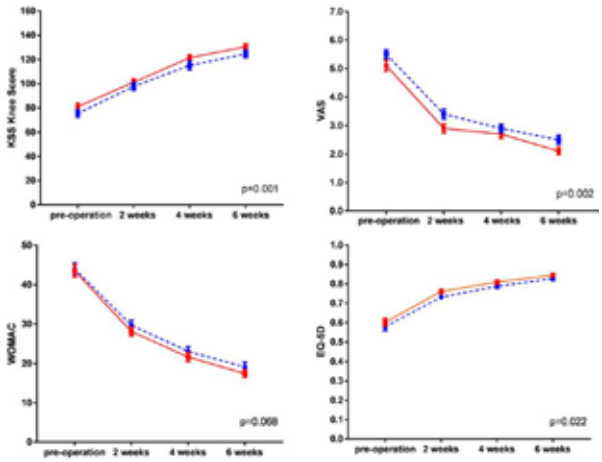


Figure 2 KSS, VAS, WOMAC and EQ-5D6 score between the two groups in the effectiveanalysis population set. The red solid lines represent the parecoxib/celecoxib group, the blue dashed lines represent the control group and error bars represent SEs calculated separately for each time point. The differences between groups were tested by linear mixed model adjusted for the gender, age, height, weight and different medical centres. KSS, Knee Society Score; VAS, Visual Analogue Scale; WOMAC, Western Ontarioand McMaster Universities Osteoarthritis Index.

use is a key component of the ERAS pathway in TKA patients.[21]

Pain control

The present study demonstrated better pain control performance and opioid-sparing effects with the sequential analgesia. Patients in the parecoxib/celecoxib group not only required less morphine, but experienced greater pain relief compared with the placebo group at all time points after surgery. Since both pain and opioid-related symptoms can hinder the patient's mobilisation and may increase the length of recovery,[21] our results suggest that the sequential analgesia regimen with parecoxib followed by celecoxib is a potential excellent choice for pain relief and enhanced recovery.

Inflammation control

Our results also showed that peripheral blood IL-6, postoperative drainage fluid PGE2, ESR and CRP were significantly decreased in the parecoxib/celecoxib group. Previous findings[22] have suggested that local inflammatory reactions triggered by tissue damage not only increase central and peripheral pain sensitivity but also lead to local intensified pain, oedema and increased bleeding at the knee joint, which is a great challenge in postoperative rehabilitation. Our present findings provide positive evidences that sequential use of COX-2 selective NSAIDs after surgery decreases surgically induced secretion of inflammatory mediators, reduces the incidence of fibrosis and the degree of local oedema, which is beneficial for postoperative exercises.

Opioid consumption

The decreased opioid consumption with the sequential analgesia provides benefits not only in reducing opioid-related adverse effects, but also in reducing overall treatment costs. Opioid drugs are associated with various dose-dependent adverse symptoms.[7 8] In the present study, we observed slightly less gastrointestinal adverse events in the parecoxib/celecoxib group (19.64%) compared with the placebo group (20.72%). Athanasakis *et al*[23] demonstrated that addition of parecoxib to opioid leads to potential savings of €858 per patient compared with opioid use alone. Therefore, reduction in the occurrence of opioid-related adverse events can lead to savings on overall treatment costs.

Adverse events

Our safety data showed that the incidence and severity of adverse events were significantly lower in the parecoxib/celecoxib group compared with the placebo group. The PRECISION trial[24] demonstrated that celecoxib at moderate doses is non-inferior to ibuprofen or naproxen with regard to cardiovascular safety. The CONCERN trial[25] concluded that in patients at high risk of both cardiovascular and gastrointestinal events, celecoxib plus proton-pump inhibitor is the preferred treatment. All these emerging evidences support the safety of selective NSAID drugs, and therefore the sequential analgesic regimen in this study.

Functional rehabilitation

This study observed significantly higher KSS and function scores of the operated knee, and better EQ-5D scores within 6 weeks post-operation in the parecoxib/celecoxib group. Theoretically, satisfactory pain management and inflammation control are beneficial to rehabilitation effectiveness, quick functional recovery and high patient satisfaction. Malan *et al*[26] and Desjardins *et al*[27] demonstrated that both parecoxib and celecoxib result in significantly improved recovery and patient satisfaction. Further well-designed trials with larger sample size and longer treatment period are suggested to elucidate the associations of NSAID use with knee function improvement and patient satisfaction after TKA.

Strengths

This study was, to our knowledge, the first randomised trial to investigate the efficacy and safety of the sequential analgesic regimen of intravenous parecoxib followed by oral celecoxib after TKA, assessing not only morphine consumption, but also pain relief, inflammation control and functional rehabilitation. In addition, compared with previous studies which attempted to observe the short-term effects of single NSAIDs, the present study showed the benefits of prolonged sequential treatment of parecoxib and celecoxib in medium-term recovery. Finally, facing the worldwide problem of opioid tolerance and related side effects, these data of the postoperative sequential regimen of NSAIDs that have been

Table 3 Cytokine, knee circumference and knee skin temperature of post operation in effectiveanalysis population set

	Parecoxib/celecoxib(n=96)	Placebo(n=97)	P value*
Postoperative drainage fluid			
PGE2			
24 hours	66.55 (27.24 to 187.66)	58.40 (21.94 to 152.09)	0.743
48 hours	600.76 (315.81 to 1022.30)	1990.64(710.50 to 5126.83)	<0.001†
72 hours	431.52 (221.37 to 819.13)	2052.73(916.46 to 4831.57)	<0.001†
Peripheral blood			
IL-6			
24 hours	3.10 (2.15 to 6.53)	3.53 (2.06 to 6.60)	0.925
48 hours	59.80 (32.85 to 105.00)	64.15 (35.50 to 131.00)	0.332
72 hours	37.50 (23.80 to 70.30)	57.45 (28.25 to 99.60)	0.009
2 weeks	4.53 (2.93 to 8.81)	7.81 (4.23 to 13.70)	0.001†
4 weeks	3.49 (2.50 to 5.64)	5.52 (3.64 to 9.12)	0.002†
6 weeks	3.50 (2.50 to 5.61)	4.08 (2.88 to 6.87)	0.177
ESR			
72 hours	42.00 (29.00 to 57.00)	62.50 (46.00 to 80.00)	<0.001†
2 weeks	28.00 (17.50 to 50.00)	49.50 (33.00 to 63.50)	<0.001†
4 weeks	20.00 (10.00 to 33.00)	28.00 (19.00 to 41.00)	0.011
6 weeks	17.00 (9.00 to 25.70)	19.00 (13.00 to 30.00)	0.028
CRP			
72 hours	78.60 (56.70 to 102.00)	117.00 (76.80 to 154.20)	<0.001†
2 weeks	8.00 (3.00 to 18.00)	14.09 (4.98 to 25.00)	0.024
4 weeks	2.85 (1.30 to 6.12)	5.39 (2.75 to 8.63)	0.006†
6 weeks	2.68 (1.43 to 5.75)	2.56 (1.35 to 5.57)	0.701
Knee circumference			
24 hours	42.53±4.15	44.22±3.67	0.003†
48 hours	43.12±4.28	44.43±3.73	0.024
72 hours	43.14±4.52	44.45±4.14	0.037
2 weeks	42.14±4.39	43.10±3.88	0.122
4 weeks	41.33±4.30	42.55±3.98	0.057
6 weeks	40.95±4.64	42.20±4.40	0.077
Knee skin temperature			
24 hours	36.62±1.30	37.05±1.19	0.017
48 hours	36.58±1.11	37.03±1.00	0.003†
72 hours	36.53±1.11	36.84±1.02	0.043
2 weeks	36.05±1.14	35.98±1.29	0.701
4 weeks	35.81±1.36	35.94±1.19	0.523
6 weeks	35.56±1.49	35.71±1.35	0.501

Data were presented as median (IQR) and tested by the independent Mann-Whitney U test.
*The significance level was corrected according to the Bonferroni correction.
†The difference was statistically significant.
CRP, C-reactive protein; ESR, erythrocyte sedation rate; IL, interleukin; PGE2, prostaglandin E2.

widely accepted as clinical routine in China, may provide important evidence to support the incorporation of this strategy into the standard multimodal analgesic regimen of the ERAS programme in OA patients undergoing TKA.

Limitations

The possible limitations of the PIPFORCE study should be mentioned. First, since the four study centres in this multicentre randomisedcontrolled trial were all in

Table 4 Adverse events between groups in the Safety set

	Parecoxib/celecoxib (n=112)	Placebo (n=111)	Absolute rate difference* (95% CI)	P value
Adverse event	25 (22.32)	45 (40.54)	−18.22 (−30.17 to −6.27)	0.003†
Severity degree				
Mild	18 (16.07)	28 (25.23)	−9.15 (−19.72 to 1.41)	0.091
Moderate	7 (6.25)	12 (10.81)	−4.56 (−11.87 to 2.75)	0.223
Serious	0 (0.00)	5 (4.50)	−4.50 (−8.36 to −0.65)	0.029†
Relationship with study treatment				
Definitely related	1 (0.89)	4 (3.60)	−2.71 (−6.59 to 1.17)	0.212
Possibly related	15 (13.39)	28 (25.23)	−11.83 (−22.08 to −1.58)	0.025†
Not related	10 (8.93)	14 (12.61)	−3.68 (−11.81 to 4.44)	0.375
Duration of AE, days	1.0 (1.0 to 2.0)	1.0 (1.0 to 3.0)	–	0.411
Expected AE	2 (1.79)	0 (0.00)	1.78 (−0.67 to 4.23)	0.498
Type of adverse events				
Gastrointestinal disorders	22 (19.64)	23 (20.72)	−1.08 (−11.61 to 9.46)	0.841
Constipation	0 (0.00)	1 (0.90)		
Diarrhoea	1 (0.89)	0 (0.00)		
Nausea	10 (8.93)	5 (4.50)		
Vomiting	11 (9.82)	17 (15.32)		
General disorders and administration site conditions	0 (0.00)	10 (9.01)	−9.01 (−14.34 to −3.68)	0.001†
Hyperhidrosis	0 (0.00)	1 (0.90)		
Pain	0 (0.00)	2 (1.80)		
Fever	0 (0.00)	7 (6.31)		
Immune system disorders	2 (1.79)	2 (1.80)	−0.02 (−3.50 to 3.47)	>0.999
Dermatitis allergic	0 (0.00)	1 (0.90)		
Drug hypersensitivity	1 (0.89)	0 (0.00)		
Hypersensitivity	1 (0.89)	1 (0.90)		
Investigations	0 (0.00)	5 (4.50)	−4.50 (−8.36 to −0.64)	0.029†
Blood glucose increased	0 (0.00)	1 (0.90)		
Body temperature increased	0 (0.00)	4 (3.60)		
Musculoskeletal and connective tissue disorders	0 (0.00)	2 (1.80)	−1.80 (−4.28 to 0.67)	0.474
Joint ankylosis	0 (0.00)	1 (0.90)		
Knee deformity	0 (0.00)	1 (0.90)		
Nervous system disorders	1 (0.89)	2 (1.80)	−0.91 (−3.93 to 2.12)	0.994
Dizziness	0 (0.00)	2 (1.80)		
Headache	1 (0.89)	0 (0.00)		
Vascular disorders	0 (0.00)	1 (0.90)	−0.90 (−2.66 to 0.86)	0.498
Venous thrombosis	0 (0.00)	1 (0.90)		

Data were presented as No. (percentage) or median (IQR) and tested by X^2 test or the independent Mann-Whitney U test.
*The absolute rate median differences were parecoxib/celecoxib group minus placebo group, and 95% CIs were calculated using Newcombe-Wilson score method.
†The difference was statistically significant.
AE, adverse event.

mainland China, the results should be interpreted with caution, and further validation studies of data sets from other institutions outside China are required. Second, the PIPFORCE study did not investigate the long-term (eg, >3 months) effects of the sequential treatment on inflammation control and functional rehabilitation after TKA. Third, although the EAP set reached 193 participants, the PP set consisted of only 155 participants (placebo group 71; parecoxib/celecoxib group 84), which is slightly less than the precalculated 172 participants according to the above sample size estimation. However, these results demonstrated significant differences in the primary outcome between the two groups in ITT, PP and EAP analyses. Furthermore, we used the PASS 14.0 software to calculate the post hoc sample size, when the cumulative opioid consumption levels until 2 weeks between the two groups were 139.3±93.6 and 58.2±44.3, respectively, only 18 participants per group could achieve a 90% power in detecting the difference between the two groups with a significance level of 0.05 (two-sided). Therefore, we believe that the present results have sufficient power to support our conclusion. Fourth, we did not conduct a time-to-event modelling analysis for AE, such as competing risk analysis. Lastly, we used general anaesthesia in this study without combining regional anaesthesia, and it should be noted that our results could not be generalised in every anaesthetic technique.

Summary

In conclusion, the PIPFORCE trial demonstrated that the sequential analgesic regimen with intravenous parecoxib followed by oral celecoxib for postsurgical analgesic treatment requires less morphine in postoperative 2 weeks. Given the increasing recognition of opioid tolerance and related side effects as well as the emerging high quality evidences for cardiovascular and gastrointestinal safety of selective NSAIDs, sequential NSAID use could play a more significant role than currently known in multimodel analgesic regimens. It should be noted that since the PIPFORCE trial was exclusively performed in mainland China, these results still require further validation studies of data sets from other institutions outside China.

Author affiliations
[1]Department of Orthopedics, Peking Union Medical College Hospital, Beijing, China
[2]Department of Orthopedics, Peking Union Medical College Hospital, Beijing, China
[3]Research Center of Clinical Epidemiology, Peking University 3rd Hospital, Beijing, China
[4]Department of Orthopedics, The Second Affiliated Hospital of Zhejiang University, Zhejiang, China
[5]Department of Orthopedics, West China Hospital, Sichuan University, Chengdu, China

Acknowledgements We would like to thank all study participants and staff who devoted their time and efforts to the study, especially to Dr Shangyi Hui (Peking Union Medical College Hospital, China), Professor Yiming Zhao (Peking University Third Hospital, China) and Miss Zhuang Liu for their critical reading and modification of the manuscript. Thanks to Caddie Qu, Jialing Yang and Math Zhang at Shanghai Bestudy Medical Technology Co, Ltd, for their statistical analysis support.

Contributors WXS contribute as the senior author and the principle investigator (PI) of this study. ZQY, as the sub PI, design the protocol, wrote the first draft of the manuscript and contributed to the coordination of the study. TLY, as the medical statistician for the study, contributed to the statistical design, acquisition and analysis of data for the work. JJ, LJ, QWW, BYY, WW, GN, FB, STZ, ZMF recruited patients for the study and participated in coordination. LYL, DYL, PHM, LY, FY were responsible for data input. LJH, YSG, SB, PFX critically revised the report. All authors reviewed the results and approved the final version of the manuscript.

Funding The PIPFORCE study was an investigator-initiated research sponsored by Pfizer (grant number: WI182703). Parecoxib sodium, celecoxib and placebo were provided by Pfizer. The funders had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript and decision to submit the manuscript for publication.

Competing interests None declared.

Patient consent for publication Obtained.

Ethics approval Approval for this study has been obtained from the Ethics Committee of Peking Union Medical College Hospital, China.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iD
Xisheng Weng http://orcid.org/0000-0003-0083-5190

REFERENCES

1 Glyn-Jones S, Palmer AJR, Agricola R, *et al*. Osteoarthritis. *The Lancet* 2015;386:376–87.
2 Carr AJ, Robertsson O, Graves S, *et al*. Knee replacement. *The Lancet* 2012;379:1331–40.
3 Thomazeau J, Rouquette A, Martinez V, *et al*. Acute pain factors predictive of post-operative pain and opioid requirement in multimodal analgesia following knee replacement. *Eur J Pain* 2016;20:822–32.
4 Lee YS. Comprehensive analysis of pain management after total knee arthroplasty. *Knee Surg Relat Res* 2017;29:80–6.
5 Gaffney CJ, Pelt CE, Gililland JM, *et al*. Perioperative pain management in hip and knee arthroplasty. *Orthop Clin North Am* 2017;48:407–19.
6 Turnbull ZA, Sastow D, Giambrone GP, *et al*. Anesthesia for the patient undergoing total knee replacement: current status and future prospects. *Local Reg Anesth* 2017;10:1–7.
7 Khademi H, Kamangar F, Brennan P, *et al*. Opioid therapy and its side effects: a review. *Arch Iran Med* 2016;19:870–6.
8 Nicholson BD. Economic and clinical burden of opioid-induced nausea and vomiting. *Postgrad Med* 2017;129:111–7.
9 Karlsen APH, Wetterslev M, Hansen SE, *et al*. Postoperative pain treatment after total knee arthroplasty: a systematic review. *PLoS One* 2017;12:e0173107.
10 Moucha CS, Weiser MC, Levin EJ. Current strategies in anesthesia and analgesia for total knee arthroplasty. *J Am Acad Orthop Surg* 2016;24:60–73.
11 Zhuang Q, Bian Y, Wang W, *et al*. Efficacy and safety of postoperative intravenous parecoxib sodium followed by oral celecoxib (PIPFORCE) post-total knee arthroplasty in patients with osteoarthritis: a study protocol for a multicentre, double-blind, parallel-group trial. *BMJ Open* 2016;6:e011732.
12 Symonds T, Hughes B, Liao S, *et al*. Validation of the Chinese Western Ontario and McMaster universities osteoarthritis index in patients from mainland China with osteoarthritis of the knee. *Arthritis Care Res* 2015;67:1553–60.
13 Liu D, He X, Zheng W, *et al*. Translation and validation of the simplified Chinese new knee Society scoring system. *BMC Musculoskelet Disord* 2015;16:391.
14 Wang H-M, Patrick DL, Edwards TC, *et al*. Validation of the EQ-5D in a general population sample in urban China. *Qual Life Res* 2012;21:155–60.

15 Nussmeier NA, Whelton AA, Brown MT, *et al*. Safety and efficacy of the cyclooxygenase-2 inhibitors parecoxib and valdecoxib after noncardiac surgery. *Anesthesiology* 2006;104:518–26.
16 Jakobsen JC, Gluud C, Wetterslev J, *et al*. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Med Res Methodol* 2017;17:162.
17 Newcombe RG. Confidence intervals for the mean of a variable taking the values 0,1 and 2. *Stat Med* 2003;22:2737–50.
18 Basse L, Hjort Jakobsen D, Billesbølle P, *et al*. A clinical pathway to accelerate recovery after colonic resection. *Ann Surg* 2000;232:51–7.
19 Ding J, Sun B, Song P, *et al*. The application of enhanced recovery after surgery (ERAS)/fast-track surgery in gastrectomy for gastric cancer: a systematic review and meta-analysis. *Oncotarget* 2017;8.
20 Mohsina S, Shanmugam D, Sureshkumar S, *et al*. Adapted eras pathway vs. standard care in patients with perforated duodenal ulcer-a randomized controlled trial. *J Gastrointest Surg* 2018;22.
21 Beverly A, Kaye AD, Ljungqvist O, *et al*. Essential elements of multimodal analgesia in enhanced recovery after surgery (ERAS) guidelines. *Anesthesiol Clin* 2017;35:e115–43.
22 Honsawek S, Deepaisarnsakul B, Tanavalee A, *et al*. Relationship of serum IL-6, C-reactive protein, erythrocyte sedimentation rate, and knee skin temperature after total knee arthroplasty: a prospective study. *Int Orthop* 2011;35:31–5.
23 Athanasakis K, Petrakis I, Vitsou E, *et al*. A cost-effectiveness analysis of parecoxib in the management of postoperative pain in the Greek health care setting. *Clin Ther* 2013;35:1118–24.
24 Nissen SE, Yeomans ND, Solomon DH, *et al*. Cardiovascular safety of celecoxib, naproxen, or ibuprofen for arthritis. *N Engl J Med* 2016;375:2519–29.
25 Chan FKL, Ching JYL, Tse YK, *et al*. Gastrointestinal safety of celecoxib versus naproxen in patients with cardiothrombotic diseases and arthritis after upper gastrointestinal bleeding (concern): an industry-independent, double-blind, double-dummy, randomised trial. *The Lancet* 2017;389:2375–82.
26 Malan TP, Marsh G, Hakki SI, *et al*. Parecoxib sodium, a parenteral cyclooxygenase 2 selective inhibitor, improves morphine analgesia and is opioid-sparing following total hip arthroplasty. *Anesthesiology* 2003;98:950–6.
27 Desjardins PJ, Grossman EH, Kuss ME, *et al*. The injectable cyclooxygenase-2-specific inhibitor parecoxib sodium has analgesic efficacy when administered preoperatively. *Anesth Analg* 2001;93:721–7.

BMJ Open

Does breast reduction surgery improve health-related quality of life? A prospective cohort study in Australian women

Tamara Crittenden ,[1,2] David I Watson,[2] Julie Ratcliffe ,[3] Philip A Griffin,[1] Nicola R Dean ,[1,2] on behalf of the AFESA Research Group

To cite: Crittenden T, Watson DI, Ratcliffe J, *et al*. Does breast reduction surgery improve health-related quality of life? A prospective cohort study in Australian women. *BMJ Open* 2020;**10**:e031804. doi:10.1136/bmjopen-2019-031804

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-031804).

Received 20 May 2019
Revised 11 December 2019
Accepted 18 December 2019

[1]Department of Plastic and Reconstructive Surgery, Flinders Medical Centre, Bedford Park, South Australia, Australia
[2]Discipline of Surgery, College of Medicine and Public Health, Flinders University, Bedford Park, South Australia, Australia
[3]Caring Futures Institute, College of Nursing and Health Sciences, Flinders University, Adelaide, South Australia, Australia

Correspondence to
Dr Nicola R Dean;
nicola.dean@sa.gov.au

ABSTRACT

Objectives To assess the health burden of breast hypertrophy and the comparative effectiveness of breast reduction surgery in improving health-related quality of life.

Design Prospective cohort study.

Setting A major public tertiary care hospital in Australia.

Participants Women with symptomatic breast hypertrophy who underwent breast reduction surgery were followed for 12 months. A comparison control cohort comprised women with breast hypertrophy who did not undergo surgery.

Interventions Bilateral breast reduction surgery for women in the surgical cohort.

Main outcome measures The primary outcome measure was health-related quality of life measured preoperatively and at 3, 6 and 12 months postoperatively using the Short Form-36 (SF-36) questionnaire. Secondary outcome measures included post-surgical complications.

Results 209 patients in the surgical cohort completed questionnaires before and after surgery. 124 patients in the control hypertrophy cohort completed baseline and 12-month follow-up questionnaires. At baseline, both groups had significantly lower scores compared with population norms across all scales ($p<0.001$). In the surgical cohort significant improvements were seen across all eight SF-36 scales ($p<0.001$) following surgery. Within 3 months of surgery scores were equivalent to those of the normal population and this improvement was sustained at 12 months. SF-36 physical and mental component scores both significantly improved following surgery, with a mean change of 10.2 and 9.2 points, respectively ($p<0.001$). In contrast, SF-36 scores for breast hypertrophy controls remained at baseline across 12 months. The improvement in quality of life was independent of breast resection weight and body mass index.

Conclusion Breast reduction significantly improved quality of life in women with breast hypertrophy. This increase was most pronounced within 3 months of surgery and sustained at 12-month follow-up. This improvement in quality of life is comparable to other widely accepted surgical procedures. Furthermore, women benefit from surgery regardless of factors including body mass index and resection weight.

Strengths and limitations of this study

- ► This large prospective longitudinal study reports 12-month follow-up using a validated patient-reported outcome measure for health-related quality of life assessment.
- ► The completion rate of the study was 83% for participants who underwent surgery.
- ► Comparisons were made with a control cohort of women with breast hypertrophy not undergoing surgery, and also to a normative female reference population.
- ► This was a non-randomised study design.

INTRODUCTION

Breast reduction surgery is a common plastic surgery procedure and it has previously been shown to be effective for relieving pain and functional problems associated with breast hypertrophy,[1–5] whereas conservative approaches to treatment such as physiotherapy, hormonal therapy and weight loss have much less impact.[6 7] However, despite clear published evidence to the contrary, breast reduction surgery is often regarded more as a cosmetic rather than a functional procedure by the general public and many medical professionals.[1 8 9] This is in spite of the finding that breast hypertrophy is a chronic health problem and relief of physical symptoms is the primary motivator for most women who are pursuing breast reduction surgery.[10]

The increasing demand for breast reduction surgery and increasing pressure to constrain healthcare spending have led to lengthy waiting times and restrictions placed on surgery in numerous countries and jurisdictions worldwide.[4 11–15] While 'rationing' of healthcare is an essential process in public healthcare systems globally, it has the potential to threaten equity of access to surgical

treatment. Within the Australian public hospital system, access to breast reduction surgery for patients is ultimately reliant on state and local policies.[16–21] Similarly, in the UK, reports on the rationing of surgery by the National Health Service (NHS) on the basis of geographical location have resulted in a 'postcode lottery'.[22 23] In 2018, reports from the NHS England 'Evidence-Based Interventions Programme' proposed to restrict funding for procedures it considers 'unnecessary', to save money and eliminate unwarranted clinical variation.[24] The inclusion of breast reduction surgery as a 'procedure of limited effectiveness' implies that it is a marginal and low priority procedure in comparison to other medical interventions.[25] However, labelling breast reduction surgery an 'ineffective' and 'unnecessary' procedure might be misleading and inaccurate, with little evidence to support this claim. Furthermore, restrictive access policies are in place in both public and private sectors in many countries and jurisdictions worldwide; often these restrictions are based on body mass index (BMI) or a minimum weight of breast resection at surgery.[4 8 9 11–15 22 23 25] The validity of such criteria might not be evidence-based, resulting in women with a medical need for surgery being denied access to it.

The primary aim of this study was to longitudinally assess health-related quality of life (HRQoL) in women with breast hypertrophy before and after breast reduction surgery, and to compare these outcomes to control groups of women with breast hypertrophy not undergoing surgery, and also to a normative female reference population. The Short Form-36 (SF-36) is a well-established indicator of patient-reported outcome for evaluating the burden of disease states and the outcomes of medical interventions and was therefore chosen as the primary outcome measure for this study. Second, this study aimed to assess the impact of patient demographics and surgical characteristics including, but not limited to, those commonly used as selection criteria for access to surgery and insurance coverage on preoperative HRQoL scores and the long-term improvement in HRQoL following surgery.

METHODS

Design and participants

A prospective cohort study was performed at Flinders Medical Centre in Adelaide, Australia. All women aged 18 years and over with symptomatic breast hypertrophy who were assessed for bilateral breast reduction surgery between April 2007 and February 2018 were informed of the study. Patients who underwent breast reduction surgery comprised the surgery cohort. Patients who were referred for surgery and were placed on the waiting list but were not expected to undergo surgery within 12 months comprised the controls.

All participants who consented to the study were asked to complete the SF-36 questionnaire at set time points. For the surgical patients this was preoperatively and 3, 6 and 12 months postoperatively. For the control patients, the questionnaire was completed at baseline and again 12 months after enrolment. Data including age, height, weight, bra cup size, health status and smoking status were determined for all patients at baseline and again at follow-up. Women who were unable to complete written questionnaires or were enrolled in the control group and had breast reduction surgery within 12 months of enrolment, or who did not return study questionnaires, were excluded from the study.

Outcome measures

The SF-36 V.2 was used to measure the general HRQoL.[26] This contains 36 items which assess health across eight subscales. Questionnaire responses were transformed as per the SF-36 V.2 scoring manual to provide the eight subscales, each with a score between 0 and 100, with higher scores indicating better health.[27] The subscales were converted into two summary scores: Physical Component Summary (PCS) score and Mental Component Summary (MCS) score using norm-based methods and scoring coefficients from the Australian population.[28] For comparison purposes, general female population normative scores were obtained from the 2008 South Australian Health Omnibus Survey and scores weighted to correspond to the age distribution of the study participants.[29]

Sample size was determined a priori and a minimum sample size of 98 patients per group was calculated to give 80% power at a two-sided significance level of 5% to detect a mean difference of 10-points with an estimated SD of 25-points in the SF-36 questionnaire score.

Study-specific questionnaires, which asked about time off work and consumption and expenditure on medications, were administered at the baseline and 12-month postoperative time points. Participants in the surgical cohort were asked postoperatively whether they would have the surgery again if they had their time over. Additional data were collected pertaining to the surgical technique used, and the weight of breast tissue removed. Hospital records were used to determine the length of hospital stay, number of outpatient clinic appointments relating to the surgery and complications leading to re-hospitalisation, or a further operative procedure within the 12 months follow-up period. A comprehensive complications checklist was completed prospectively during the study by the treating doctor. Three-dimensional laser body scanning was performed preoperatively and at 12 months postoperatively using a Cyberware WBX scanner (Cyberware) and Cyslice software (Headus Pty Ltd). Breast and body volume were measured from the scan according to a protocol described previously.[30 31]

Statistical analysis

Statistical analyses were performed using SPSS V.25.0 statistical software (IBM Corp). Descriptive statistics including mean, SD and 95% CI were computed for continuous variables. Comparisons between groups were

made using t-tests for continuous data and χ^2 tests for categorical data, with Fisher's exact test as appropriate. Linear mixed models were used to assess the significance of changes in SF-36 subscale scores over multiple time points. For each SF-36 scale an improvement score was calculated using the score obtained at the last available assessment, with a higher score representing a greater improvement from baseline. Pearson correlation coefficients were calculated to assess the linear association between SF-36 scores and baseline participant and clinical characteristics; variables that showed a significant association were entered into the regression model. Candidate variables included age, BMI, preoperative breast volume, bra cup size, tissue resection weight (grams), breast asymmetry and ratio of breast to body volume. Variables were continuous except for bra cup size which was categorised into six groups as follows: D, DD, E, F, G and ≥H cup. Multiple linear regression was used to assess whether any of the collected sociodemographic or clinical variables were associated with first, SF-36 PCS score at baseline, and second, with the change in SF-36 PCS scores from baseline to 12 months after surgery. Statistical significance was accepted at a p value of less than 0.05.

Patient and public involvement

At the design stage of the study two group meetings were held with women with breast hypertrophy to discuss their perspective on the condition, deliver education material and discuss this study. In addition, one consumer was more extensively involved with the design of the study and trialling different methods of breast volume measurement. Study results will be disseminated to the public through presentations and local health newsletter.

Table 1 Baseline characteristics of participants

Characteristic	Surgical cohort (n=209)	Hypertrophy control cohort (n=124)	P value of difference*
Mean (SD; range) age (years)	42.6 (13.4; 18 to 72)	45.3 (13.1; 20 to 79)	0.079
Age group (years):			
18–24	24 (12)	12 (10)	
25–34	38 (18)	15 (12)	
35–44	64 (31)	26 (21)	
45–54	41 (20)	43 (34)	
55–64	31 (15)	21 (18)	
≥65	11 (5)	7 (6)	
Mean (SD) BMI (kg/m^2)	32.7 (6.0)	32.2 (6.1)	0.468
Obesity status:			
Non-obese (<30)	71 (34)	48 (39)	0.326
Obese (≥30)	138 (66)	74 (61)	
Missing	0 (0)	2 (0)	
Smoking status:			
Non-smoker	108 (52)	78 (63)	0.243
Current smoker	35 (17)	14 (11)	
Ex-smoker <12 months	15 (7)	5 (5)	
Ex-smoker >12 months	47 (23)	25 (20)	
Missing	4 (0)	2 (0)	
Bra cup size:			
≤D	13 (6)	4 (3)	
DD	43 (21)	13 (11)	
E	50 (24)	19 (15)	
F	46 (22)	27 (22)	
G	35 (17)	37 (30)	
≥H	19 (10)	19 (15)	
Missing	3 (0)	5 (0)	

Values are numbers (percentages) unless stated otherwise.
*Using independent samples t-test or χ^2 test as appropriate.
BMI, body mass index.

RESULTS

Surgical cohort

Of 251 participants who completed a baseline assessment and underwent bilateral breast reduction surgery, 209 (83.3%) completed at least one postoperative follow-up assessment and were included in the study group for analysis. Missing data were due to participants repeatedly not attending appointments or choosing to not complete and return the study questionnaires at some time points. Twenty-three participants formally withdrew from the study following surgical intervention. Baseline characteristics were compared between participants who were lost to follow-up and those who completed at least one postoperative assessment. No difference was observed for age, BMI, tissue weight resected or preoperative SF-36 scales and summary scores except for the mental health scale, where non-respondents had a lower mean score of 6.8 points less than responders (p=0.034).

Participant demographics for the surgical cohort are summarised in table 1. Preoperatively, mean total breast volume measured by 3D laser scanner was 3391 mL (range 1472–9622 mL). At 12 months postoperatively, mean total breast volume was 2184 mL (range 963 to 4392 mL). The mean total weight of breast tissue resected at surgery was 1338 g±817 g. An inferior pedicle breast reduction technique was the most commonly used approach (161/209, 77%), followed by a superior pedicle technique (35/209, 17%). The average hospital stay was 2.3 days. Fifty-nine patients (28%) experienced at least one surgical complication. Eight patients (3.8%) had subsequent procedures for revision of surgical scars or to correct 'dog-ears'.

The majority of participants (204/209, 97.6%) responded in the postoperative questionnaire that they would have the surgery again, while others were either unsure (4/209, 1.9%) or would not have surgery again (1/209, 0.5%). Following surgery, participants on average spent less money on medications and treatments (AU$26.41 vs AU$5.73 per month, p<0.001) and took fewer days off work (4.5 days vs 0.1 days in the previous 6-month period, p=0.009) when compared with before surgery. Using bivariate analysis, obesity was not associated with an increased incidence of surgical complications (p=0.323), with the incidence of complications in non-obese participants (17/71, 24%) and obese participants (42/138, 30%). Furthermore, there were no differences in the incidence of major complications based on obesity status.

The SF-36 was completed preoperatively and at least once postoperatively by 209 surgical participants; 191 (91%) completed the postoperative questionnaires at 3 months, 183 (88%) at 6 months and 193 (92%) at 12 months. When compared with previously published age-adjusted normative data for the female Australian population,[29] mean baseline SF-36 scores for the surgical cohort were significantly lower across all scales (p<0.001) (table 2). A comparison of mean preoperative and 3-month postoperative SF-36 scores showed that scores were significantly higher across all eight SF-36 subscales (p<0.001) (table 2) such that they reached the level of the normative population (figure 1). Mean SF-36 PCS and MCS scores significantly improved following surgery, increasing by 10.2 (95% CI; 8.2 to 12.1) and 9.2 (95% CI; 6.9 to 11.6) points, respectively (p<0.001) (figure 2 and online supplementary table S1). The mean change in SF-36 PCS and MCS scores was in excess of the developer-recommended 3-point minimal important difference (MID) threshold.[32 33] SF-36 scores were stable at 6 and 12 months post-surgery and linear mixed-model analysis showed no significant difference from those at 3 months post-surgery. The mean change in SF-36 scores from baseline to 12 months following surgery was in excess of MID threshold estimates based on a rule of thumb 10-point change on 100-point quality of life scales[34] or 0.5 SD default value for patient-perceived important change[35] in all eight SF-36 subscales (figure 2). SF-36 scores for obese women improved equally, if not greater than their non-obese counterparts following surgery, reaching statistical significance for the physical functioning subscale (table 3).

Breast hypertrophy control cohort

Study questionnaires were initially posted to 350 women with breast hypertrophy who were not scheduled for surgery; 160 (46%) completed and returned the questionnaires at baseline, and of these 124 responded again 12 months later. Twenty-four of those contacted to participate in the study underwent breast reduction surgery during the study time frame and were therefore excluded. Participant demographics for the hypertrophy control cohort are summarised in table 1. No significant differences were observed when comparing spending on medications and number of days off work between baseline and 12 months following enrolment, with both remaining significantly higher than postoperative surgical participants (p<0.001).

Mean baseline SF-36 scores for women in the breast hypertrophy control group were significantly lower than the normative population across all dimensions (table 2). At 12 months post-baseline, SF-36 scores showed no significant improvement and remained significantly lower than population norms (table 2) and postoperative scores for women in the surgical cohort (figure 2). Mean SF-36 PCS and MCS summary scores for women in the breast hypertrophy control group were significantly lower than those who underwent breast reduction surgery, with a mean difference of 10.6 (95% CI; 8.3 to 12.8) and 11.1 points (95% CI; 8.2 to 13.9), respectively (p<0.001) (table 2).

Comparing the improvement in HRQoL with other surgical interventions

The improvement in SF-36 physical and mental summary scores in women who underwent surgery in our study was compared with existing studies which describe 12-month postoperative outcomes from other surgical interventions

Table 2 Mean (95% CI) SF-36 scores for participants in the surgical cohort, hypertrophy control cohort and normative female population

SF-36 scale	Normative* (n=1551)	Hypertrophy control cohort		Surgical cohort			
		Baseline (n=160)	12 months (n=124)	Preoperative (n=209)	3 months postoperative (n=190)	6 months postoperative (n=181)	12 months postoperative (n=191)
PF	84.2 (83.2 to 85.2)	64.7 (60.9 to 68.6)	61.1 (56.7 to 65.7)	61.0 (57.6 to 64.5)	80.1 (76.9 to 83.3)	80.8 (77.3 to 84.4)	83.4 (80.2 to 86.5)
RP	82.0 (80.7 to 83.3)	58.3 (53.8 to 62.8)	58.1 (53.0 to 63.0)	56.0 (52.2 to 59.9)	79.5 (76.1 to 82.9)	81.1 (77.4 to 84.8)	81.3 (77.7 to 84.9)
BP	73.0 (71.9 to 74.1)	39.8 (36.3 to 40.4)	37.9 (34.2 to 41.7)	38.5 (35.6 to 41.5)	67.4 (63.9 to 70.8)	67.6 (63.6 to 71.7)	71.6 (67.9 to 75.2)
GH	70.2 (69.1 to 71.4)	49.7 (46.3 to 53.1)	49.8 (46.0 to 54.0)	57.9 (55.0 to 60.9)	69.1 (66.3 to 72.0)	69.5 (66.6 to 72.4)	70.4 (67.7 to 73.1)
VT	57.3 (56.2 to 58.3)	36.7 (33.6 to 39.8)	35.1 (31.3 to 38.8)	39.7 (37.0 to 42.5)	57.7 (54.8 to 60.6)	58.6 (55.8 to 61.3)	58.9 (56.1 to 61.7)
SF	82.6 (81.3 to 83.8)	55.2 (50.6 to 59.8)	55.1 (50.1 to 59.9)	57.1 (53.3 to 60.9)	78.8 (75.0 to 82.5)	79.4 (75.6 to 83.2)	81.4 (78.1 to 84.7)
RE	88.3 (87.2 to 89.3)	62.8 (58.0 to 67.5)	60.2 (55.4 to 65.7)	61.7 (57.8 to 65.7)	80.1 (76.5 to 83.7)	82.3 (78.9 to 85.7)	84.6 (81.4 to 87.7)
MH	77.0 (76.1 to 78.0)	58.8 (55.2 to 62.5)	56.1 (52.4 to 59.9)	59.8 (57.0 to 62.5)	73.7 (71.0 to 76.4)	73.8 (71.1 to 76.5)	74.3 (71.6 to 76.9)
PCS	49.7 (49.2 to 50.2)	39.6 (38.1 to 41.1)	39.3 (37.5 to 41.1)	39.7 (38.4 to 41.0)	48.9 (47.6 to 50.3)	49.0 (47.5 to 50.5)	49.9 (48.4 to 51.3)
MCS	47.6 (47.0 to 48.2)	36.2 (33.8 to 38.6)	35.1 (32.7 to 37.6)	37.0 (35.2 to 38.8)	45.4 (43.6 to 47.1)	45.7 (44.0 to 47.4)	46.2 (44.5 to 47.9)

*Source: age-standardised normative data from the South Australian female population.[29]
BP, Bodily pain; GH, General health; MCS, Mental Component Summary; MH, Mental health; PCS, Physical Component Summary; PF, Physical function; RE, Role emotional; RP, Role physical; SF-36, Short Form-36; SF, Social function; VT, Vitality.

(table 4). Breast reduction surgery provided a greater gain in SF-36 PCS scores than a coronary artery bypass graft and hernia repair and the improvement was similar to that experienced by patients undergoing total knee replacement surgery. The improvement in SF-36 MCS scores following breast reduction surgery exceeded that of all other surgical procedures.

The impact of participant characteristics on HRQoL and benefit of surgical intervention

There was a significant positive correlation between baseline BMI and the total amount of breast tissue resected at surgery. That is, as the BMI increased there was an associated increase in the amount of breast tissue removed (Pearson's r=0.641, p<0.001). When exploring baseline SF-36 PCS scores, a significant negative correlation was found between SF-36 PCS scores and age (r=−0.13), BMI (r=−0.30), tissue resection weight (r=−0.26), degree of breast hypertrophy (r=−0.28) and ratio of breast to body volume (r=−0.19). Multivariate linear regression of candidate variables against baseline SF-36 PCS scores found BMI to be the only variable significantly related to preoperative SF-36 PCS scores (R^2=0.16, p<0.001). Multivariate regression analysis was also used to analyse predictors of the change in SF-36 PCS score following surgery and showed that improvement in SF-36 PCS scores was not significantly associated with any of these factors.

DISCUSSION

Principal findings

Findings from this study demonstrate that women with symptomatic breast hypertrophy have impaired quality of life compared with those in the general population. At baseline, participants in both the surgical and control breast hypertrophy groups scored significantly lower than the female general population in all SF-36 subscales, with pain being the most prominent. Surgical participants quality of life improved following breast reduction to such an extent that the health deficits were eliminated at 3 months following surgery and quality of life was 'normalised' to levels equivalent to that of the general population across all dimensions. This normalisation effect was stable across 12 months follow-up. The SF-36 health gain ranged from 14.5 to 33.1 points, and this exceeded the minimally important difference threshold estimates of one-half a SD approach[35] or a rule-of-thumb of a 10-point change on 100-point subscales,[34] supporting the contention that breast reduction surgery provides a clinically relevant health benefit.

Secondary aims of this study were to investigate factors that have the potential to influence the level of improvement in quality of life following surgery: BMI, degree of hypertrophy, bra cup size, age, preoperative breast symmetry and weight of tissue resection at surgery. Several of these factors are frequently used to restrict access to breast reduction surgery, none of which are based on high-quality evidence. In our study the improvement in

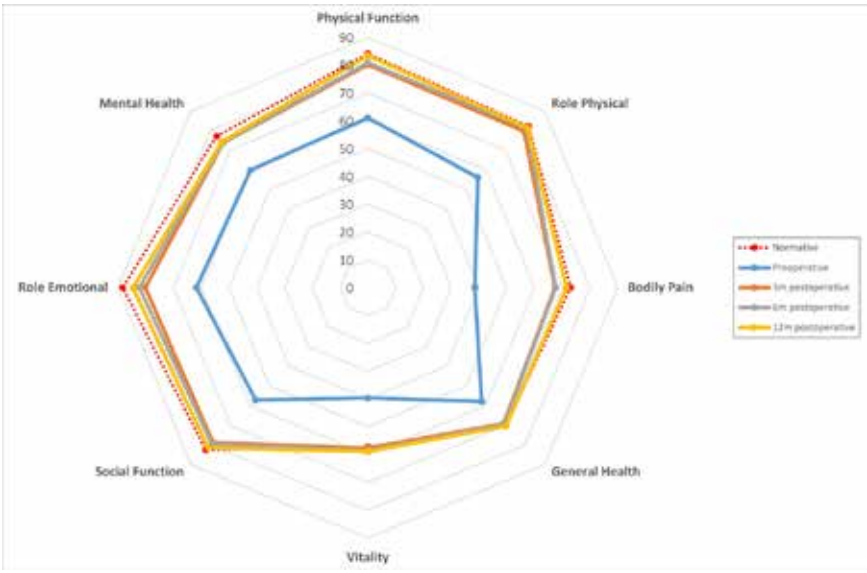


Figure 1 Comparison of mean preoperative and postoperative Short Form-36 scores with age-standardised female population norms (South Australian Health Omnibus Survey).[29]

HRQoL was independent of these factors, suggesting that all women with symptomatic breast hypertrophy can benefit from this surgery regardless of commonly scrutinised factors. This is of clinical relevance as it highlights that women with a higher BMI or those with a lower weight of resection benefit equally and should not be discriminated against based on criteria-based restrictions. Furthermore, there was no increase in the complication rate in the obese participants.

Comparison with other studies

The finding that women with symptomatic breast hypertrophy have a considerable health deficit and impaired quality of life compared with women in the general population is supported by existing studies within the literature.[4 14 36–39] These studies also report that surgical intervention provides symptomatic relief and improves HRQoL to levels of the general population. Our findings support those of Blomqvist *et al* and demonstrate that the improvement in quality of life is stable for up to 1 year after surgery, enabling women to return to levels of HRQoL that are similar to the normal population.[1]

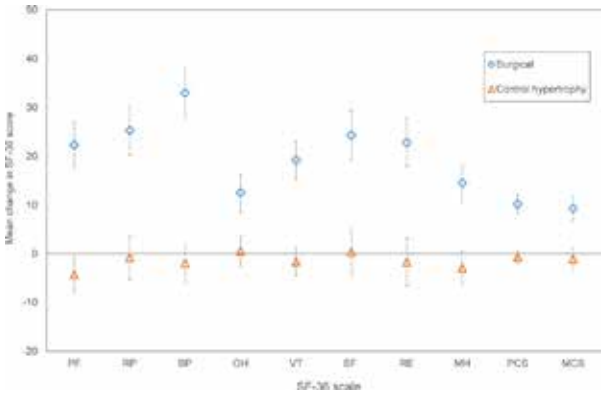


Figure 2 Mean change in Short Form-36 (SF-36) scores from baseline to 12 months for surgical and breast hypertrophy control groups. Error bars represent 95% CI. BP, bodily pain; GH, general health; MCS, mental component summary; MH, mental health; PCS, physical component summary; PF, physical function; RE, role emotional; RP, role physical; SF, social function, VT, vitality.

Our study demonstrated that symptom relief and improvement in HRQoL are not impacted by BMI or the removal of a minimum weight of resection. This finding is consistent with existing studies using the SF-36; however, two of these studies were potentially biased due to the BMI restrictions on their study populations.[6 40 41] Our study also supports previous findings of no significant difference in the complication rate based on obesity status.[41–43] In spite of these findings access restrictions for breast reduction surgery are in place in many countries, despite a lack of supporting evidence.

The intervention effect of breast reduction surgery in our study was well in excess of the minimal clinically important difference for SF-36 PCS and MCS scores, which has been recommended by the developers as a 3-point change.[32 33] The improvements in the SF-36 PCS score at 1 year following surgery were comparable to those of other widely accepted surgical interventions such as total hip and total knee replacement,[44] spinal fusion,[45] bariatric surgery[46] and coronary artery bypass graft surgery.[47] The improvements in the mental component score following breast reduction surgery actually exceeded those of all other interventions cited. Breast reduction surgery is a relatively inexpensive procedure, and the improvement in HRQoL provides evidence as to the comparative effectiveness of this intervention in

Table 3 Comparison of mean change (95% CI) in SF-36 scores following surgery in non-obese and obese participants

SF-36 subscale	Non-obese (n=71)	Obese (n=138)	Difference in means (95% CI)	P value of difference*
Physical function	16.1 (11.2 to 22.1)	23.4 (19.5 to 27.3)	6.8 (0.03 to 13.5)	0.050
Role physical	19.4 (12.4 to 26.3)	25.9 (21.2 to 30.5)	6.5 (–1.7 to 14.7)	0.121
Bodily pain	28.6 (22.8 to 34.5)	32.3 (27.8 to 36.9)	3.7 (–3.9 to 11.4)	0.337
General health	10.2 (6.0 to 14.3)	12.2 (8.4 to 16.0)	2.0 (–4.1 to 8.2)	0.516
Vitality	18.9 (14.8 to 23.1)	18.3 (14.2 to 22.4)	–0.7 (–7.2 to 5.9)	0.842
Social function	23.6 (17.8 to 29.4)	21.9 (16.6 to 27.2)	–1.7 (–10.2 to 6.8)	0.701
Role emotional	18.9 (12.8 to 25.0)	22.5 (17.2 to 27.8)	3.6 (–5.0 to 12.2)	0.409
Mental health	14.9 (11.4 to 18.5)	13.0 (9.2 to 16.9)	–1.9 (–7.9 to 4.1)	0.532

Obesity status: non-obese (<30 kg/m^2), obese (≥30 kg/m^2).
*Using an independent t-test.
SF-36, Short Form-36.

relieving the health burden and the functional symptoms of breast hypertrophy.

Strengths and limitations of this study

A potential limitation of our study was that the participant response rate for the breast hypertrophy control cohort was relatively low at 46%, which may be due to the recruitment process via postal questionnaire. Furthermore, while the total follow-up period for this cohort was 12 months, the intermediate time points of 3 and 6 months that were collected in the surgical cohort were not included in this cohort, although the consistency of outcomes at baseline and 12 months suggest that 3 and 6 month outcomes are likely to have been similar.

The strengths of our study were the prospective design, the relatively large sample size and the inclusion of a non-surgical control sample of women with breast hypertrophy who were recruited from the same waiting list as those in the surgical cohort. In addition, the postoperative outcomes described in this study included multiple time points over a 12-month period. In addition, our surgical cohort was not biased by restrictions that have been reported in previous studies based on a minimum weight of resection or BMI and therefore includes a broad spectrum across these variables. This is particularly important as it enables the accurate assessment of these factors as potential predictors of the change in HRQoL and outcomes of surgery and overcomes these limitations.

CONCLUSIONS AND POLICY IMPLICATIONS

Breast hypertrophy is a painful condition which is effectively treated by breast reduction surgery. The marked improvement in quality of life following breast reduction surgery is comparable to other widely accepted and approved surgical interventions. This study highlights that the improvement in quality of life following surgery is independent of traditionally used criteria based on BMI or a minimum weight of resection and demonstrates the health benefits of surgery regardless of these factors. This confirms the clinical effectiveness of breast reduction surgery and supports the contention that there is no justification for excluding women based on criteria such as BMI or the extent of breast hypertrophy.

Acknowledgements We thank the participants for their involvement in the study, especially Diane Barker for her valuable contribution; Andrea Smallman, Suzanne Patten and Richard Woodman for statistical advice; and Anne Taylor, Tiffany Gill and

Table 4 Mean improvement in SF-36 PCS and MCS scores between surgical interventions

Reference	Surgical intervention	Preop PCS	Postop PCS	ΔPCS	Preop MCS	Postop MCS	ΔMCS	N
This study	Bilateral breast reduction	39.7	49.9	10.2	37.0	46.2	9.2	191
Pivec *et al*[44]	Total knee replacement	33.0	47.8	14.8	52.9	55.9	3.0	281
Stickles *et al*[48]	Total hip replacement	28.0	41.2	13.2	51.2	53.9	2.7	551
Muller-Nordhorn *et al*[47]	Coronary artery bypass grafting	36.0	43.0	7.3	45.0	50.0	4.3	412
Polly *et al*[45]	Lumbar fusion (spine)	26.6	40.0	13.4	n/a	n/a	n/a	1826
Rogmark *et al*[49]	Incisional hernia repair	41.6	49.5	8.1	50.2	52.3	1.7	124
Faulconbridge *et al*[46]	Bariatric surgery	37.7	46.4	8.7	43.1	45.5	2.4	36

Δ, mean change in SF-36 score from preoperative to 12 months postoperative; MCS, Mental Component Summary; N, number of participants; n/a, not applicable; PCS, Physical Component Summary; SF-36, Short Form-36.

Jodie Avery for providing normative Short Form-36 data and scoring algorithms. TC is the recipient of an Australian Government Research Training Programme PhD Scholarship.

Collaborators The following are members of the Anthropometry for Elective Surgery Assessment (AFESA) Research Group: Tamara Crittenden, Nicola Dean (Flinders Medical Centre, Flinders University); Philip Griffin, Karen Burford, Kaveeta Deut, Elizabeth Lomax, Phillipa van Essen (Flinders Medical Centre); Maciej Henneberg (University of Adelaide) and Daisy Veitch (SHARP Dummies, Flinders University).

Contributors NRD, TC, PAG, JR, DIW, DV and MH contributed to study design. TC, KB, EL, PE and KD contributed to data collection. TC was responsible for statistical analyses and original preparation of the paper. TC, JR, DIW and NRD drafted the manuscript. All authors approved the final manuscript and contributed intellectually important content. NRD attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. NRD is the guarantor of the paper and takes full responsibility for the work and/or the conduct of the study, had access to the data and controlled the decision to publish.

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 None declared.

Patient consent for publication Not required.

Ethics approval This study was approved by the Southern Adelaide Clinical Human Research Ethics Committee (SAC HREC reference numbers 118.056 and 73.17).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No additional data available.

ORCID iDs
Tamara Crittenden http://orcid.org/0000-0002-6484-3977
Julie Ratcliffe http://orcid.org/0000-0001-7365-1988
Nicola R Dean http://orcid.org/0000-0001-7084-2359

REFERENCES

1 Blomqvist L, Eriksson A, Brandberg Y. Reduction mammaplasty provides long-term improvement in health status and quality of life. *Plast Reconstr Surg* 2000;106:991–7.
2 Freire M, Neto MS, Garcia EB, *et al*. Functional capacity and postural pain outcomes after reduction mammaplasty. *Plast Reconstr Surg* 2007;119:1149–56. discussion 57-8.
3 Mello AA, Domingos NAM, Miyazaki MC. Improvement in quality of life and self-esteem after breast reduction surgery. *Aesthetic Plast Surg* 2010;34:59–64.
4 Klassen A, Fitzpatrick R, Jenkinson C, *et al*. Should breast reduction surgery be rationed? A comparison of the health status of patients before and after treatment: postal questionnaire survey. *BMJ* 1996;313:454–7.
5 Thoma A, Sprague S, Veltri K, *et al*. A prospective study of patients undergoing breast reduction surgery: health-related quality of life and clinical outcomes. *Plast Reconstr Surg* 2007;120:13–26.
6 Collins ED, Kerrigan CL, Kim M, *et al*. The effectiveness of surgical and nonsurgical interventions in relieving the symptoms of macromastia. *Plast Reconstr Surg* 2002;109:1556–66.
7 Iwuagwu OC, Walker LG, Stanley PW, *et al*. Randomized clinical trial examining psychosocial and quality of life benefits of bilateral breast reduction surgery. *Br J Surg* 2006;93:291–4.
8 Frey JD, Koltz PF, Bell DE, *et al*. The complex insurance reimbursement landscape in reduction mammaplasty: how does the American plastic surgeon navigate it? *Ann Plast Surg* 2014;72:23–9.
9 Taylor AJ, Tate D, Brandberg Y, *et al*. Cost-Effectiveness of reduction mammaplasty. *Int J Technol Assess Health Care* 2004;20:269–73.
10 Davis GM, Ringler SL, Short K, *et al*. Reduction mammaplasty: long-term efficacy, morbidity, and patient satisfaction. *Plast Reconstr Surg* 1995;96:1106–10.
11 Breuning EE, Oikonomou D, Singh P, *et al*. Cosmetic surgery in the NHS: applying local and national guidelines. *J Plast Reconstr Aesthet Surg* 2010;63:1437–42.
12 Koltz PF, Frey JD, Langstein HN. Insurance coverage and reduction mammaplasty: a systematic review of current health care policies. *Plast Reconstr Surg* 2013;132:692e–3.
13 Nguyen JT, Wheatley MJ, Schnur PL, *et al*. Reduction mammaplasty: a review of managed care medical policy coverage criteria. *Plast Reconstr Surg* 2008;121:1092–100.
14 O'Blenes CAE, Delbridge CL, Miller BJ, *et al*. Prospective study of outcomes after reduction mammaplasty: long-term follow-up. *Plast Reconstr Surg* 2006;117:351–8.
15 Seitchik MW. Reduction mammaplasty: criteria for insurance coverage. *Plast Reconstr Surg* 1995;95:1029–32.
16 ACT Government. Waiting time and elective surgery access policy 2016 January 25, 2018. Available: https://www.act.gov.au/search?query=access+to+elective+surgery&collection=act-gov&profile=www-act
17 NSW Government Department of Health. Waiting time and elective surgery policy, 2012. Available: http://www1.health.nsw.gov.au/pds/ActivePDSDocuments/PD2012_011.pdf
18 Government of South Australia SA Health. Restricted elective surgery policy directive, 2018. Available: https://www.sahealth.sa.gov.au/wps/wcm/connect/3d568a77-a030-437a-957e-9b60bb143341/Directive_Restricted+Elective+Surgery_03072018.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-3d568a77-a030-437a-957e-9b60bb143341-mhygh9U
19 Tasmanian Government Department of Health and Human Services. Improving Time to Treatment - Tasmania's Elective Surgery Access Policy, 2015. Available: http://www.dhhs.tas.gov.au/__data/assets/pdf_file/0020/60176/Improving_Time_to_Treatment_-_Tasmania_s_Elective_Surgery_Access_-_DHHS_Wide_-_THOs_-_Policy_-_200911201_-_UNDER_REVIEW.pdf
20 Victorian Government Department of Human Services. Elective surgery access policy, 2015. Available: https://www2.health.vic.gov.au/about/publications/policiesandguidelines/elective-surgery-access-policy-2015
21 Government of Western Australia. Elective surgery access and waiting list management policy, 2017. Available: https://ww2.health.wa.gov.au/About-us/Policy-frameworks/Clinical-Services-Planning-and-Programs/Mandatory-requirements/Elective-Services/Elective-Surgery-Access-and-Waiting-List-Management-Policy
22 Royal College of Surgeons of England. Is access to surgery a Postcode lottery? 2014. Available: https://www.rcseng.ac.uk/-/media/files/rcs/news-and-events/is-access-to-surgery-a-postcode-lottery.pdf?la=en
23 Wraight WM, Tay SKL, Nduka C, *et al*. Bilateral breast reduction surgery in England: a postcode lottery. *J Plast Reconstr Aesthet Surg* 2007;60:1039–44.
24 Iacobucci G. NHS proposes to stop funding 17 "unnecessary" procedures. *BMJ* 2018;362:k2903.
25 NHS England. Evidence-Based interventions: response to the public consultation and next steps, 2018. Available: https://www.england.nhs.uk/wp-content/uploads/2018/11/ebi-consultation-response-document.pdf
26 Ware JE, Sherbourne CD. The mos 36-item short-form health survey (SF-36). I. conceptual framework and item selection. *Med Care* 1992;30:473–83.
27 Ware J, Kosinski MA, Dewey J. *How to score version 2 of the SF-36® health survey*, 2000.
28 Hawthorne G, Osborne RH, Taylor A, *et al*. The SF36 version 2: critical analyses of population weights, scoring algorithms and population norms. *Qual Life Res* 2007;16:661–73.
29 Marin T, Taylor A, Gill T. *A population profile of quality of life in South Australia—population norms for 2008 using the SF36 version 2. Population research and outcome studies unit. South Australian*. Adelaide: Department of Health, 2009.
30 Veitch D, Burford K, Dench P, *et al*. Measurement of breast volume using body scan technology(computer-aided anthropometry). *Work* 2012;41:4038–45.
31 Yip JM, Mouratova N, Jeffery RM, *et al*. Accurate assessment of breast volume: a study comparing the volumetric gold standard (direct water displacement measurement of mastectomy specimen) with a 3D laser scanning technique. *Ann Plast Surg* 2012;68:135–41.
32 Frendl DM, Ware JE. Patient-Reported functional health and well-being outcomes with drug therapy: a systematic review of randomized trials using the SF-36 health survey. *Med Care* 2014;52:439–45.
33 Ware JE, Kosinski M, Bayliss MS, *et al*. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and

summary measures: summary of results from the medical outcomes study. *Med Care* 1995;33:AS264–79.
34 King MT. The interpretation of scores from the EORTC quality of life questionnaire QLQ-C30. *Qual Life Res* 1996;5:555–67.
35 Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;41:582–92.
36 Miller BJ, Morris SF, Sigurdson LL, *et al*. Prospective study of outcomes after reduction mammaplasty. *Plast Reconstr Surg* 2005;115:1025–31.
37 Shakespeare V, Cole RP. Measuring patient-based outcomes in a plastic surgery service: breast reduction surgical patients. *Br J Plast Surg* 1997;50:242–8.
38 Kececi Y, Sir E, Gungor M. Patient-Reported quality-of-life outcomes of breast reduction evaluated with generic questionnaires and the breast reduction assessed severity scale. *Aesthet Surg J* 2015;35:48–54.
39 Blomqvist L, Brandberg Y. Three-Year follow-up on clinical symptoms and health-related quality of life after reduction mammaplasty. *Plast Reconstr Surg* 2004;114:49–54.
40 Eggert E, Schuss R, Edsander-Nord A. Clinical outcome, quality of life, patients' satisfaction, and aesthetic results, after reduction mammaplasty. *Scand J Plast Reconstr Surg Hand Surg* 2009;43:201–6.
41 Güemes A, Pérez E, Sousa R, *et al*. Quality of life and alleviation of symptoms after breast reduction for Macromastia in obese patients: is surgery worth it? *Aesthetic Plast Surg* 2016;40:62–70.
42 Wagner DS, Alfonso DR. The influence of obesity and volume of resection on success in reduction mammaplasty: an outcomes study. *Plast Reconstr Surg* 2005;115:1034–8.
43 Cunningham BL, Gear AJL, Kerrigan CL, *et al*. Analysis of breast reduction complications derived from the BRAVO study. *Plast Reconstr Surg* 2005;115:1597–604.
44 Pivec R, Issa K, Given K, *et al*. A prospective, longitudinal study of patient satisfaction following total knee arthroplasty using the short-form 36 (SF-36) survey stratified by various demographic and comorbid factors. *J Arthroplasty* 2015;30:374–8.
45 Polly DW, Glassman SD, Schwender JD, *et al*. Sf-36 PCS benefit-cost ratio of lumbar fusion comparison to other surgical interventions: a thought experiment. *Spine* 2007;32:S20–6.
46 Faulconbridge LF, Wadden TA, Thomas JG, *et al*. Changes in depression and quality of life in obese individuals with binge eating disorder: bariatric surgery versus lifestyle modification. *Surg Obes Relat Dis* 2013;9:790–6.
47 Müller-Nordhorn J, Roll S, Willich SN. Comparison of the short form (SF)-12 health status instrument with the SF-36 in patients with coronary heart disease. *Heart* 2004;90:523–7.
48 Stickles B, Phillips L, Brox WT, *et al*. Defining the relationship between obesity and total joint arthroplasty. *Obes Res* 2001;9:219–23.
49 Rogmark P, Petersson U, Bringman S, *et al*. Quality of life and surgical outcome 1 year after open and laparoscopic incisional hernia repair: PROLOVE: a randomized controlled trial. *Ann Surg* 2016;263:244–50.

SPECIAL ARTICLE

Introduction of Surgical Safety Checklists in Ontario, Canada

David R. Urbach, M.D., Anand Govindarajan, M.D., Refik Saskin, M.Sc., Andrew S. Wilton, M.Sc., and Nancy N. Baxter, M.D., Ph.D.

ABSTRACT

BACKGROUND

Evidence from observational studies that the use of surgical safety checklists results in striking improvements in surgical outcomes led to the rapid adoption of such checklists worldwide. However, the effect of mandatory adoption of surgical safety checklists is unclear. A policy encouraging the universal adoption of checklists by hospitals in Ontario, Canada, provided a natural experiment to assess the effectiveness of checklists in typical practice settings.

METHODS

We surveyed all acute care hospitals in Ontario to determine when surgical safety checklists were adopted. Using administrative health data, we compared operative mortality, rate of surgical complications, length of hospital stay, and rates of hospital readmission and emergency department visits within 30 days after discharge among patients undergoing a variety of surgical procedures before and after adoption of a checklist.

RESULTS

During 3-month periods before and after adoption of a surgical safety checklist, a total of 101 hospitals performed 109,341 and 106,370 procedures, respectively. The adjusted risk of death during a hospital stay or within 30 days after surgery was 0.71% (95% confidence interval [CI], 0.66 to 0.76) before implementation of a surgical checklist and 0.65% (95% CI, 0.60 to 0.70) afterward (odds ratio, 0.91; 95% CI, 0.80 to 1.03; P=0.13). The adjusted risk of surgical complications was 3.86% (95% CI, 3.76 to 3.96) before implementation and 3.82% (95% CI, 3.71 to 3.92) afterward (odds ratio, 0.97; 95% CI, 0.90 to 1.03; P=0.29).

CONCLUSIONS

Implementation of surgical safety checklists in Ontario, Canada, was not associated with significant reductions in operative mortality or complications. (Funded by the Canadian Institutes of Health Research.)

From the Institute for Clinical Evaluative Sciences (D.R.U., A.G., R.S., A.S.W., N.N.B.), the Department of Surgery (D.R.U., A.G., N.N.B.) and Institute of Health Policy, Management and Evaluation (D.R.U., N.N.B.), University of Toronto, the University Health Network (D.R.U.), Mount Sinai Hospital (A.G.), and Keenan Research Centre, Li Ka Shing Knowledge Institute, Department of Surgery, St. Michael's Hospital (N.N.B.) — all in Toronto. Address reprint requests to Dr. Urbach at 200 Elizabeth St., 10-214, Toronto, ON M5G 2C4, Canada, or at david.urbach@uhn.ca.

Drs. Urbach and Baxter contributed equally to this article.

N Engl J Med 2014;370:1029-38.
DOI: 10.1056/NEJMsa1308261

A study published in 2009 showed that implementation of the 19-item World Health Organization (WHO) Surgical Safety Checklist substantially reduced the rate of surgical complications, from 11.0% to 7.0%, and reduced the rate of in-hospital death from 1.5% to 0.8%.[1] The WHO estimated that at least 500,000 deaths per year could be prevented through worldwide implementation of this checklist.[2] This dramatic effect of a relatively simple and accessible intervention resulted in its widespread adoption. In the United Kingdom, a nationwide program was implemented by the National Health Service within weeks after publication of the WHO study,[3] and almost 6000 hospitals worldwide are actively using or have expressed interest in using the checklist.[4]

The effect of mandatory checklist implementation is unclear. Studies of implementation have been observational,[5-11] have been limited to a small number of centers,[6-11] have not evaluated patient outcomes,[8-10] or have not shown the magnitude of effectiveness found in the WHO study.[6,7] Only studies including team training[11-13] or a more comprehensive safety system that includes multiple checklists[14] have shown effectiveness similar to that seen in the WHO study.

Implementation of surgical safety checklists is not uniform,[15,16] and performance quality may be lower when participation is not voluntary. In Ontario, a Canadian province with a population of more than 13 million people, the Ministry of Health and Long-Term Care mandated public reporting of adherence to surgical safety checklists for hospitals beginning in July 2010.[17] The rapid implementation of surgical safety checklists in Ontario provided a natural experiment to evaluate the effectiveness of checklist implementation at the population level.

METHODS

OVERVIEW

We analyzed the outcomes of surgical procedures performed before and after the adoption of surgical safety checklists, using population-based administrative health data (see the Supplementary Appendix, available with the full text of this article at NEJM.org). The study was approved by the research ethics board of Sunnybrook Health Sciences Centre.

SURGICAL SAFETY CHECKLISTS

We contacted all 133 surgical hospitals in Ontario to determine when the surgical safety checklist was introduced (the month, if the day was not known), whether a special intervention or educational program was used, and the specific checklist used (the Canadian Patient Safety Institute checklist, the WHO checklist, or a unique checklist devised by the hospital). Hospitals were required to report the number of surgical procedures for which a surgical safety checklist was used (numerator) as a proportion of the total number of surgical procedures performed (denominator) at the institution. Hospitals typically designate a checklist coordinator, often an operating-room nurse, to determine whether the checklist is completed for each surgical procedure performed.[18] Compliance with surgical safety checklists is reported publicly by the Ontario Ministry of Health and Long-Term Care at the level of the individual hospital.[19]

STUDY PERIODS

We studied 3-month intervals for each hospital, one ending 3 months before the introduction of a surgical checklist, and one starting 3 months after the introduction of the checklist. We conducted sensitivity analyses using different periods for comparison.

SURGICAL PROCEDURES

We included all surgical procedures performed during each study interval. Procedure types (see the Supplementary Appendix) were selected on the basis of Canadian Classification of Health Interventions codes.[20] Some patients underwent more than one surgical procedure in one or both periods; we limited the analysis to the first procedure per patient in each study interval.

OUTCOMES

Operative mortality, defined as the rate of death occurring in the hospital or within 30 days after surgery regardless of place, was the primary outcome. We used administrative data to assess the rates of complications occurring within 30 days after surgery (see the Supplementary Appendix). We also assessed length of hospital stay, rates of readmission within 30 days after discharge, and rates of emergency department visits within 30 days after discharge.

COVARIATES

We measured comorbidity using the resource utilization bands (simplified morbidity categories) of the Adjusted Clinical Group system (0, nonusers; 1, healthy users; 2, users with low morbidity; 3, users with moderate morbidity; 4, users with high morbidity; and 5, users with very high morbidity),[21] age (0 to 17, 18 to 39, 40 to 64, and 65 years of age or older), sex, urban or rural residence, and quintile of median neighborhood household income (an ecologic measure of socioeconomic status). We also assessed attributes of the surgical intervention: admission category (ambulatory or inpatient), procedure status (emergency or elective), and month performed.

STATISTICAL ANALYSIS

In analyses of the effect of checklists on surgical outcomes, we used generalized estimating equations to adjust for potentially confounding variables and to account for the clustering of observations within hospitals.[22] We used Poisson generalized-estimating-equation models to estimate length of stay for inpatient procedures and binomial (logistic-regression) models for other outcomes. Adjusted risks were estimated with the use of the average value of each adjustment variable in the study population (age, sex, procedure status [emergency vs. elective], admission category [inpatient vs. ambulatory], urban vs. rural residence, procedure type, month of surgery, and comorbidity score). To explore associations between other variables and surgical outcomes, we also conducted analyses with adjustment for all these factors as well as for the patient's neighborhood income quintile. Since generalized-estimating-equation models did not converge for some of the infrequent surgical outcomes, we used generalized linear models to estimate the effect of checklists on surgical outcomes in analyses of specific surgical complications.

For each hospital, we estimated the age-, sex-, and month-adjusted changes in operative mortality, risk of surgical complications, length of stay, and risk of readmission or emergency department visit and plotted these values with 95% confidence intervals. The effect of the checklist did not vary substantially according to the type of checklist used (Table S1 in the Supplementary Appendix). To determine whether enthusiasm for using checklists was associated with effect, we tested interactions between the date of checklist adoption and the effect on surgical outcomes, under the assumption that earlier adopters of checklists had greater enthusiasm for their use. A priori, we planned five subgroup analyses to explore the effect of the introduction of a surgical safety checklist in subgroups defined by age, sex, procedure status, admission category, and procedure type. To test whether the effect of the checklist varied according to subgroup, we fit a separate generalized linear model for each subgroup analysis, with an interaction term specifying the joint effect of the checklist and the subgroup categories, adjusting for all other subgroup variables except those defining the subgroup analysis. All reported P values are two-sided. P values lower than 0.05 were considered to indicate statistical significance.

RESULTS

HOSPITALS AND CHECKLISTS

We retrieved information on the use of surgical safety checklists from 130 of 133 hospitals listed by the Ministry of Health and Long-Term Care as providing surgical services. Some hospitals did not perform procedures during the study period, and some multisite hospitals introduced the checklist at the same time at all sites and had a single hospital identifier, which left 101 hospitals suitable for analysis. All hospitals introduced a surgical safety checklist between June 2008 and September 2010. More than a third of the hospitals (37) began using a checklist in April 2010. Ninety-two of the 101 hospitals provided copies of their checklist; 79 used a Canadian Patient Safety Institute version (see the Supplementary Appendix), 9 used customized checklists, and 4 used the WHO checklist. Ninety-seven hospitals used a special intervention or educational program for checklist implementation. Hospital-reported compliance with checklists was high. Almost all of the 97 large community hospitals reported compliance of 99% or 100% during the period from January through June 2013. The lowest reported compliance by a large community hospital during this period was 91.6%.[19]

The number of surgical procedures performed per hospital ranged from 9 to 4422 (median, 654) during the 3-month interval before the checklist was implemented and from 2 to 4522 (median, 633)

during the 3-month interval after implementation. During both periods, nearly 90% of procedures were elective, and nearly 40% were performed during inpatient hospitalizations (Table 1, and Table S2 in the Supplementary Appendix).

EFFECT OF INTRODUCTION OF CHECKLISTS

The adjusted risk of death in the hospital or within 30 days after discharge was 0.71% (95% confidence interval [CI], 0.66 to 0.76) before and 0.65% (95% CI, 0.60 to 0.70) after implementation of a surgical safety checklist (P=0.07) (Table 2). There was a significant but small and clinically unimportant decrease in the adjusted length of stay, from 5.11 days (95% CI, 5.08 to 5.14) before checklist introduction to 5.07 days (95% CI, 5.04 to 5.10) afterward (P=0.003). There was no significant improvement in the adjusted risk of an emergency department visit within 30 days after discharge (10.44% [95% CI, 10.26 to 10.62] before implementation and 10.55% [95% CI, 10.37 to 10.73] afterward, P=0.37) or of readmission (3.11% [95% CI, 3.01 to 3.22] and 3.14% [95% CI, 3.03 to 3.24], respectively; P=0.76).

The adjusted risk of surgical complications within 30 days after the procedure was 3.86% (95% CI, 3.76 to 3.96) before implementation of a checklist and 3.82% (95% CI, 3.71 to 3.92) afterward (P=0.53). The risks of most complications did not differ significantly between the two periods. The only complication for which the risk significantly decreased was an unplanned return to the operating room (from 1.94% [95% CI, 1.87 to 2.00] to 1.78% [95% CI, 1.72 to 1.85], P=0.001). After introduction of a checklist, there were increases in the adjusted risk of deep venous thrombosis (from 0.03% [95% CI, 0.02 to 0.05] to 0.07% [95% CI, 0.05 to 0.08], P<0.001) and ventilator use (from 0.08% [95% CI, 0.06 to 0.10] to 0.12% [95% CI, 0.10 to 0.14], P=0.007).

In additional regression analyses of other determinants of surgical outcomes that also included adjustment for income quintile, the results of checklist introduction were similar. Introduction of a checklist was associated with an odds ratio of 0.91 (95% CI, 0.80 to 1.03) for operative mortality (P=0.13) and 0.97 (95% CI, 0.80 to 1.03) for surgical complications (P=0.29) (see Table S3 in the Supplementary Appendix).

EFFECT OF CHECKLISTS IN INDIVIDUAL HOSPITALS

Figure 1 shows the effect of introducing surgical safety checklists in individual hospitals. No hospital had a significant change in operative mortality after checklist introduction (Fig. 1A). Within-hospital changes in other surgical outcomes were mixed (Fig. 1B, and Fig. S1A, S1B, and S1C in the Supplementary Appendix). For example, six hospitals had significantly fewer complications after introduction of a checklist, whereas three had significantly more complications (Fig. 1B).

SUBGROUP ANALYSES

The effect of checklists did not vary substantially according to date of adoption (before, around, or after April 2010) (Table S1 in the Supplementary Appendix), which suggests that there was no benefit conferred by earlier versus later adoption. Stratified analyses did not reveal any subgroup with a significant reduction in operative mortality associated with introduction of a surgical safety checklist (Fig. 2A). There was no significant reduction in operative mortality associated with checklist introduction among subgroups at higher risk for operative death, such as persons undergoing emergency procedures (4.51% [95% CI, 4.16 to 4.86] before introduction and 4.12% [95% CI, 3.77 to 4.46] afterward, P=0.11) or inpatient procedures (1.71% [95% CI, 1.59 to 1.83] and 1.58% [95% CI, 1.46 to 1.69], respectively; P=0.11). For surgical complications (Fig. 2B), we found interactions between checklist introduction and both procedure type and admission category, with a significant increase in risk associated with checklist use for ambulatory procedures (odds ratio, 2.55; 95% CI, 1.61 to 4.03) but no significant effect for inpatient procedures (odds ratio, 0.97; 95% CI, 0.92 to 1.02; P<0.001 for interaction). The effect of the checklist on length of hospital stay differed for elective and emergency procedures and among some procedure types (Fig. S2A in the Supplementary Appendix). There were no differences among subgroups in the effect of surgical checklist introduction on the risk of readmission (Fig. S2B in the Supplementary Appendix). The results of sensitivity analyses testing longer and shorter intervals before and after checklist introduction were similar to the results of primary analyses.

DISCUSSION

In contrast to other studies, our population-based study of surgical safety checklists in Ontario hospitals showed no significant reduction in operative mortality after checklist implementation. Adjusted operative mortality was 0.71% before

Table 1. Characteristics of the Patients.*

Characteristic	Before Checklist Introduction (N = 109,341)	After Checklist Introduction (N = 106,370)
	number (percent)	
Procedure status		
Elective	97,040 (88.7)	93,699 (88.1)
Emergency	12,301 (11.3)	12,671 (11.9)
Admission category		
Ambulatory	66,660 (61.0)	64,718 (60.8)
Inpatient	42,681 (39.0)	41,652 (39.2)
Procedure type†		
Eye	21,578 (19.7)	21,471 (20.2)
Orocraniofacial	9,663 (8.8)	9,582 (9.0)
Digestive	12,867 (11.8)	13,206 (12.4)
Genitourinary	17,785 (16.3)	16,340 (15.4)
Musculoskeletal	31,381 (28.7)	30,554 (28.7)
Other	9,855 (9.0)	9,410 (8.8)
Age		
0–17 yr	7,689 (7.0)	7,806 (7.3)
18–39 yr	18,955 (17.3)	18,232 (17.1)
40–64 yr	43,669 (39.9)	42,023 (39.5)
≥65 yr	39,028 (35.7)	38,309 (36.0)
Sex		
Female	63,591 (58.2)	61,672 (58.0)
Male	45,750 (41.8)	44,698 (42.0)
Comorbidity score‡		
0–2	5,544 (5.1)	5,450 (5.1)
3	51,935 (47.5)	49,856 (46.9)
4	32,325 (29.6)	31,457 (29.6)
5	19,537 (17.9)	19,607 (18.4)
Neighborhood income quintile§		
Unknown	406 (0.4)	414 (0.4)
1	19,574 (17.9)	19,098 (18.0)
2	21,223 (19.4)	20,684 (19.4)
3	22,078 (20.2)	21,216 (19.9)
4	23,392 (21.4)	22,698 (21.3)
5	22,668 (20.7)	22,260 (20.9)
Hospital type¶		
Community	77,026 (70.4)	74,817 (70.3)
Pediatric	1,808 (1.7)	1,827 (1.7)
Small	1,713 (1.6)	1,690 (1.6)
Teaching	28,794 (26.3)	28,002 (26.3)

* Percentages may not sum to 100 because of rounding. Table S2 in the Supplementary Appendix provides a complete description of patient characteristics. Each study period was 3 months long, extending from 6 months to 3 months before checklist introduction and from 3 months to 6 months after checklist introduction.

† Categories are from the Canadian Classification of Interventions. The "other" category includes procedures involving the nervous system, respiratory system, cardiovascular system, lymphatic system, and ear.

‡ Comorbidity was assessed as the resource utilization band, a component of a six-level simplified morbidity categorization in the Adjusted Clinical Groups system[21]; it is defined by health resource use, with 0 indicating nonusers and 5 indicating users with very high morbidity.

§ Neighborhood income quintiles were calculated for the median household income in the neighborhood of a patient's residence; 1 denotes the lowest income category, and 5 the highest.

¶ Small hospitals, as defined by the Joint Policy and Planning Commission of the Ontario Ministry of Health and Long-Term Care, are hospitals with fewer than 50 inpatient beds and a referral population of fewer than 20,000 residents. Community hospitals are nonteaching hospitals.

Table 2. Surgical Outcomes before and after Introduction of a Surgical Safety Checklist.*

Outcome	Before Checklist Introduction	After Checklist Introduction	P Value†
Rate of death in the hospital or within 30 days after discharge — % (95% CI)			
Unadjusted	0.70 (0.65–0.75)	0.66 (0.61–0.71)	0.27
Adjusted	0.71 (0.66–0.76)	0.65 (0.60–0.70)	0.07
Length of hospital stay — days (95% CI)‡			
Unadjusted	5.07 (5.01–5.13)	5.11 (5.05–5.17)	0.02
Adjusted	5.11 (5.08–5.14)	5.07 (5.04–5.10)	0.003
Rate of emergency department visit within 30 days after discharge — % (95% CI)			
Unadjusted	10.28 (10.10–10.46)	10.71 (10.52–10.90)	0.001
Adjusted	10.44 (10.26–10.62)	10.55 (10.37–10.73)	0.37
Rate of readmission within 30 days after discharge — % (95% CI)			
Unadjusted	3.08 (3.00–3.18)	3.17 (3.07–3.28)	0.21
Adjusted	3.11 (3.01–3.22)	3.14 (3.03–3.24)	0.76
Rate of complications — % (95% CI)			
Unadjusted	3.80 (3.69–3.92)	3.87 (3.76–3.99)	0.41
Adjusted	3.86 (3.76–3.96)	3.82 (3.71–3.92)	0.53
Adjusted rate of specific complications — % (95% CI)			
Acute renal failure	0.10 (0.08–0.12)	0.13 (0.11–0.15)	0.08
Bleeding	0.64 (0.59–0.68)	0.63 (0.58–0.67)	0.76
Cardiac arrest	0.10 (0.08–0.12)	0.12 (0.10–0.14)	0.20
Coma	0.00 (0.00–0.01)	0.01 (0.00–0.01)	0.46
Deep venous thrombosis	0.03 (0.02–0.05)	0.07 (0.05–0.08)	<0.001
Acute myocardial infarction	0.29 (0.26–0.32)	0.29 (0.26–0.32)	0.91
Ventilator use	0.08 (0.06–0.10)	0.12 (0.10–0.14)	0.007
Pneumonia	0.31 (0.27–0.34)	0.31 (0.28–0.34)	0.80
Pulmonary embolism	0.03 (0.02–0.04)	0.03 (0.02–0.04)	0.58
Stroke	0.15 (0.12–0.17)	0.16 (0.14–0.18)	0.35
Major disruption of wound	0.14 (0.12–0.16)	0.13 (0.11–0.16)	0.61
Infection of surgical site	0.61 (0.56–0.65)	0.64 (0.59–0.69)	0.30
Sepsis	0.10 (0.08–0.11)	0.09 (0.07–0.11)	0.73
Septic shock	0.05 (0.03–0.06)	0.05 (0.04–0.06)	0.83
Unplanned return to operating room‡	1.94 (1.87–2.00)	1.78 (1.72–1.85)	0.001
Vascular graft failure	0.01 (0.00–0.02)	0.02 (0.01–0.02)	0.15
Shock	0.07 (0.06–0.09)	0.09 (0.07–0.10)	0.26

* Rates were adjusted with the use of generalized linear models for age, sex, procedure type, procedure status (emergency vs. elective), admission category (inpatient vs. ambulatory), rural or urban residence, month of surgery, and comorbidity score (assessed as the resource utilization band).
† P values are for the comparison of values before and after introduction of the checklist.
‡ The model included only inpatient hospitalizations.

and 0.65% after checklist introduction. Checklist use did not result in reductions in risks of surgical complications, emergency department visits, or hospital readmissions within 30 days after discharge. There was a significant but small and not clinically relevant reduction in adjusted length of hospital stay (5.11 days before checklist introduction and 5.07 days afterward). Surgical check-

lists did not reduce the risk of operative death in any subgroup we studied, including high-risk groups such as elderly patients, patients who underwent emergency procedures, and patients who underwent inpatient procedures.

The absence of meaningful improvements in outcomes after surgical checklist implementation was unexpected in light of the findings of studies evaluating the effects of such checklists.[1,6,11,14] In a meta-analysis of three before-and-after studies evaluating the effect of surgical safety checklists,[5] the pooled relative risk of operative death was 0.57 (95% CI, 0.42 to 0.76), and the relative risk of complications was 0.63 (95% CI, 0.58 to 0.67). Our inability to replicate these large effects cannot be explained by inadequate power; our study included more than 200,000 surgical procedures in 101 hospitals.

Ontario hospitals implemented surgical checklists between June 2008 and September 2010 in response to the plan of the Ontario Ministry of Health and Long-Term Care to publicly report compliance with use of the checklist. Self-reported compliance by all hospitals in the province is high: 92% from April through June 2010 and never less than 98% after June 2010.[19] Although materials were available to assist in the implementation of surgical safety checklists in hospitals,[23] no formal team training was required before public reporting, and implementation was not standardized. Real-world compliance with checklists varies.[24] In one hospital in the Netherlands, surgical safety checklists were fully completed for only 39% of surgical procedures after mandatory implementation.[6] In that study, the odds ratio for death in the period after implementation, as compared with the period before implementation, was reduced only among patients who underwent procedures with full checklist compliance (0.23; 95% CI, 0.16 to 0.33). There was no reduction in the odds ratio for death among patients for whom the checklist was partially completed (1.16; 95% CI, 0.95 to 1.41) or not completed (1.57; 95% CI, 1.31 to 1.89). Although selection bias probably explains much of the negative effect of noncompliance in hospitals where checklists are used, this study highlighted the fact that checklists are not always applied in a uniform manner. The absence of an effect of checklist implementation in our study may therefore reflect inadequate adherence to the checklist in Ontario. The approach to implementation in Ontario was consistent with WHO recommendations[25] and was similar to that used in many other jurisdictions.[3,26-28] It is possible that published evidence regarding the efficacy of implementing checklists within hos-

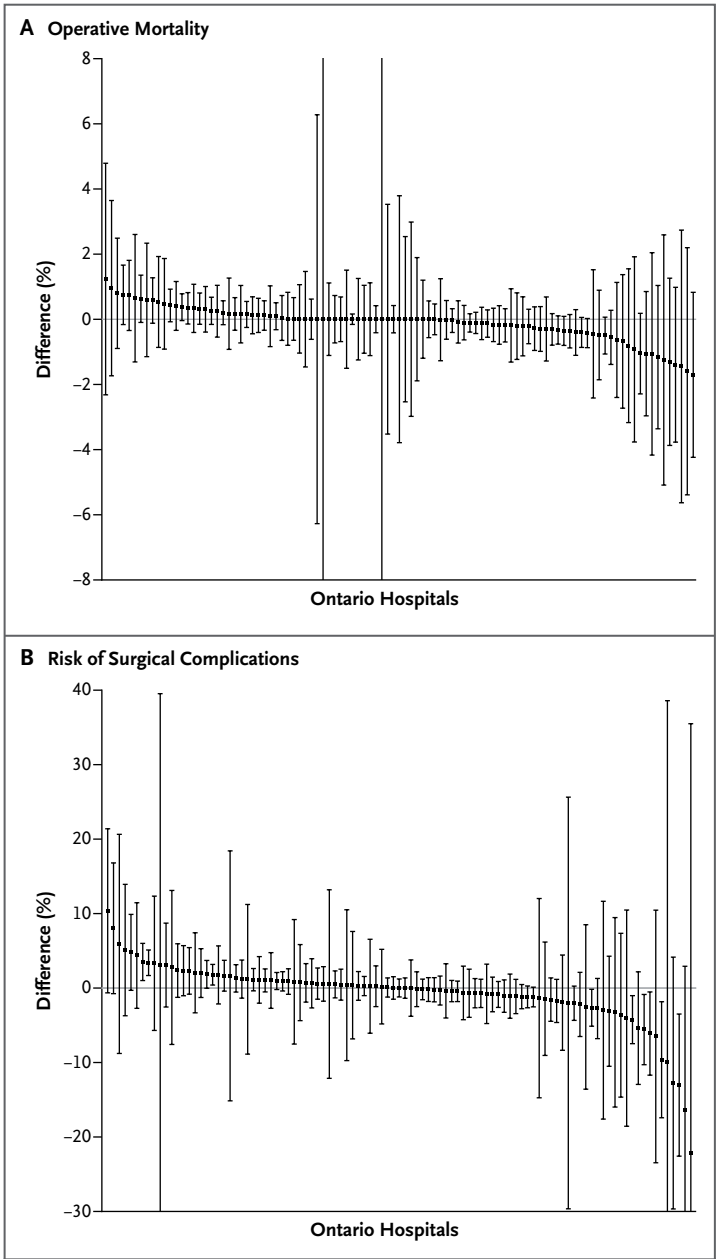


Figure 1. Within-Hospital Changes in Operative Mortality and Risk of Surgical Complications.

Each data point represents the difference in operative mortality (Panel A) and the risk of surgical complications (Panel B) before and after the implementation of a surgical safety checklist in one hospital, adjusted for age, sex, and month of surgery. Negative values indicate improvement. Hospitals are ordered from those with the highest values (least improvement) to those with the lowest values (most improvement). I bars denote 95% confidence intervals.

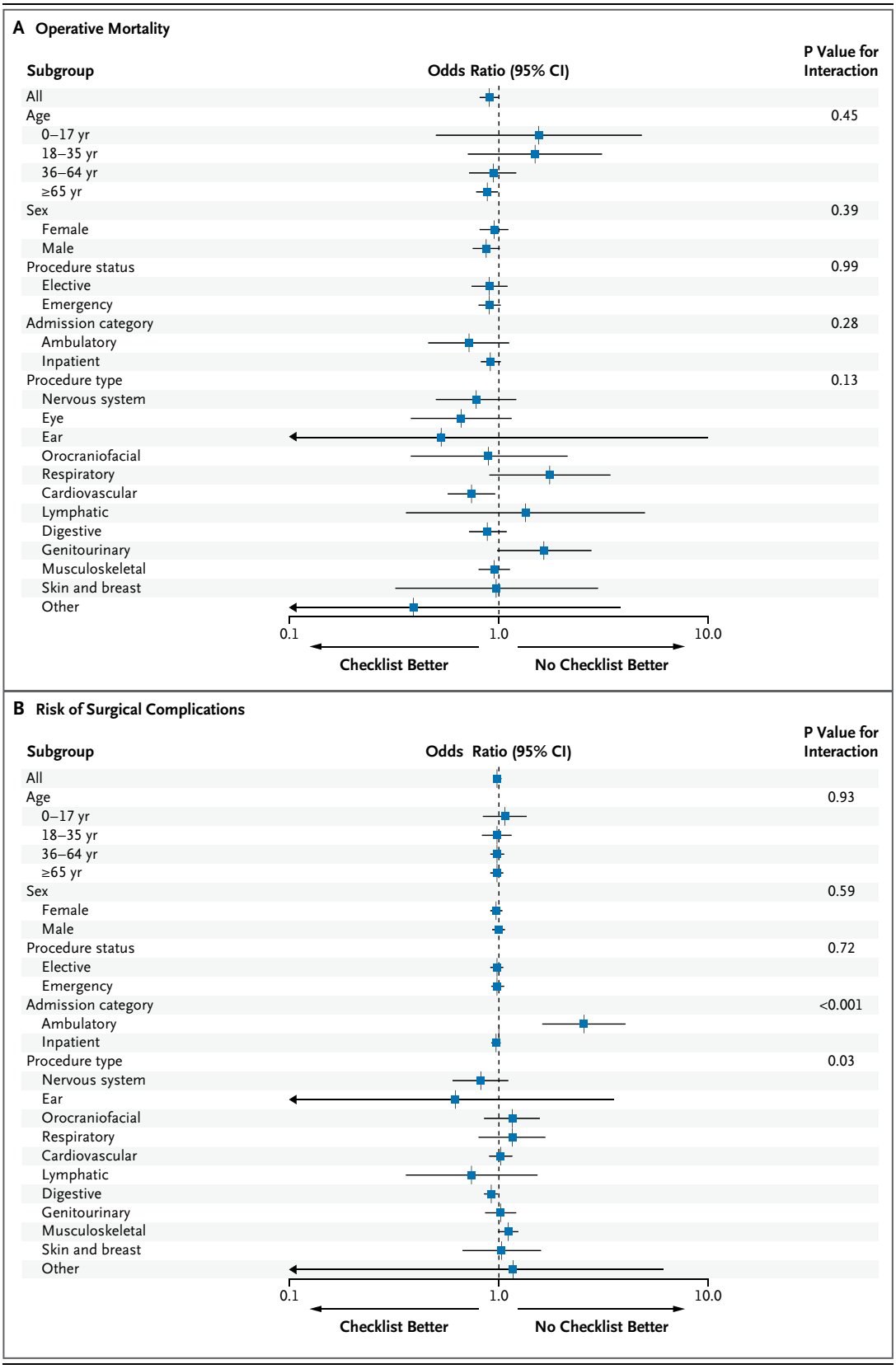


Figure 2 (facing page). Odds Ratios for Operative Mortality and Surgical Complications, Stratified According to Age, Sex, Procedure Status, Admission Category, and Type of Procedure.

Adjusted effect sizes for operative mortality (Panel A) and risk of surgical complications (Panel B) in each stratum were estimated with the use of generalized linear models, with adjustment for all variables shown except the stratification variable. For surgical complications, an odds ratio for the Eye procedure type could not be estimated because of the small number of events. P values are for the interaction between the stratification variable and the effect of checklist use on the outcome.

pitals participating in safety research is not generalizable; the effectiveness of surgical checklists in typical practice settings — as in this study — may be more limited.

It is also possible that the surgical safety checklist is less effective in practice than suggested by the existing literature. A Hawthorne effect — the tendency for some people to perform better when they perceive that their work is under scrutiny — may explain the strong effect of surgical checklists in studies in which hospitals were aware of the intervention under study. Before-and-after comparisons[1] are uncontrolled observational designs with inherent limitations, and inferences of causality should be made with caution.[29] The effectiveness of a surgical safety checklist has never been shown in a controlled trial with randomization, despite the feasibility of using cluster-randomized designs to test context-dependent interventions such as strategies for ensuring patient safety. Studies showing a substantial effect of a checklist, apart from the WHO study,[1] either coupled the checklist with extensive team training[11-13] or used an expansive checklist that covered care from the preoperative period to discharge from the hospital.[14]

In some of the 101 hospitals in this study, outcomes did change significantly — for better or worse — after implementation of a checklist. Because thousands of hospitals around the world have implemented surgical safety checklists, many will have improvements in the outcomes by chance alone. Hospital-based studies showing improvements in outcomes after checklist implementation are more likely to be published than are negative studies (publication bias[30]). The population-based nature of our study, which included virtually all hospitals providing surgical care for the population of Ontario, allowed us to obtain an estimate of the effectiveness of surgical safety checklists that is less susceptible to biases from selective reporting of institutional experience.

Our study has a number of limitations. First, secular trends and major cointerventions during the period when checklists were introduced may have confounded our results. However, we used an analytic approach similar to that used in the studies that showed a significant effect of checklists.[1,14] No other Ontario-wide interventions to improve surgical quality were implemented during the study period. Since surgical outcomes tend to improve over time,[31] it is highly unlikely that confounding due to time-dependent factors prevented us from identifying a significant improvement after implementation of a surgical checklist. Second, we used administrative data to assess surgical complications. Although this method is commonly used,[32-34] it is inferior to prospective measurement or chart review[35-37] and may have obscured changes in surgical complications after checklist implementation. However, the other outcomes studied, including operative mortality, length of stay, emergency department visits, and readmission, are less susceptible to misclassification in administrative data.

In conclusion, our study of the implementation of surgical safety checklists in Ontario did not show the striking improvement in patient outcomes identified in previous studies. We did not identify any subgroup that particularly benefited from checklists. Although a greater effect of surgical safety checklists might occur with more intensive team training or better monitoring of compliance, surgical safety checklists, as implemented during the study period, did not result in improved patient outcomes at the population level. There may be value in the use of surgical safety checklists, such as enhanced communication and teamwork and the promotion of a hospital culture in which safety is a high priority; however, these potential benefits did not translate into meaningful improvements in the outcomes we analyzed.

The opinions, results, and conclusions reported in this article are those of the authors and are independent from the funding sources. No endorsement by the Institute for Clinical Evaluative Sciences (ICES) or the Ontario Ministry of Health and Long-Term Care is intended or should be inferred.

Supported by a grant from the Canadian Institutes of Health Research (MOP-89704) and by the ICES, which is funded by an

annual grant from the Ontario Ministry of Health and Long-Term Care.

No potential conflict of interest relevant to this article was reported.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Mohammed Mamun for collecting information on checklist implementation in individual hospitals.

REFERENCES

1. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. N Engl J Med 2009;360:491-9.
2. Safe surgery saves lives. Geneva: World Health Organization (http://www.who.int/patientsafety/safesurgery/en).
3. Patient safety alert: WHO Surgical Safety Checklist. Geneva: World Health Association, 2009 (http://www.nrls.npsa.nhs.uk/resources/clinical-specialty/surgery/?entryid45=59860&p=3).
4. Patient safety: surgical safety Web map. Geneva: World Health Organization (http://maps.cga.harvard.edu:8080/Hospital/#).
5. Borchard A, Schwappach DL, Barbir A, Bezzola P. A systematic review of the effectiveness, compliance, and critical factors for implementation of safety checklists in surgery. Ann Surg 2012;256:925-33.
6. van Klei WA, Hoff RG, van Aarnhem EE, et al. Effects of the introduction of the WHO "Surgical Safety Checklist" on in-hospital mortality: a cohort study. Ann Surg 2012;255:44-9.
7. Sewell M, Adebibe M, Jayakumar P, et al. Use of the WHO Surgical Safety Checklist in trauma and orthopaedic patients. Int Orthop 2011;35:897-901.
8. Helmiö P, Blomgren K, Takala A, Pauniaho SL, Takala RS, Ikonen TS. Towards better patient safety: WHO Surgical Safety Checklist in otorhinolaryngology. Clin Otolaryngol 2011;36:242-7.
9. Takala RS, Pauniaho SL, Kotkansalo A, et al. A pilot study of the implementation of WHO surgical checklist in Finland: improvements in activities and communication. Acta Anaesthesiol Scand 2011;55:1206-14.
10. Nilsson L, Lindberget O, Gupta A, Vegfors M. Implementing a pre-operative checklist to increase patient safety: a 1-year follow-up of personnel attitudes. Acta Anaesthesiol Scand 2010;54:176-82.
11. Bliss LA, Ross-Richardson CB, Sanzari LJ, et al. Thirty-day outcomes support implementation of a surgical safety checklist. J Am Coll Surg 2012;215:766-76.
12. Young-Xu Y, Neily J, Mills PD, et al. Association between implementation of a medical team training program and surgical morbidity. Arch Surg 2011;146:1368-73.
13. Neily J, Mills PD, Young-Xu Y, et al. Association between implementation of a medical team training program and surgical mortality. JAMA 2010;304:1693-700.
14. de Vries EN, Prins HA, Crolla RM, et al. Effect of a comprehensive surgical safety system on patient outcomes. N Engl J Med 2010;363:1928-37.
15. Conley DM, Singer SJ, Edmondson L, Berry WR, Gawande AA. Effective surgical safety checklist implementation. J Am Coll Surg 2011;212:873-9.
16. Vogts N, Hannam JA, Merry AF, Mitchell SJ. Compliance and quality in administration of a Surgical Safety Checklist in a tertiary New Zealand hospital. N Z Med J 2011;124:48-58.
17. Patient Safety. Toronto: Ontario Ministry of Health and Long Term Care (http://www.health.gov.on.ca/en/public/programs/patient_safety).
18. Surgical Safety Checklist. Toronto: Ontario Hospital Association (http://www.oha.com/CurrentIssues/keyinitiatives/PatientSafety/Pages/SurgicalSafetyChecklist.aspx).
19. Public reporting — patient safety: percent of surgeries in which a surgical safety checklist was performed, April 2010–June 2013. Toronto: Health Quality Ontario (http://www.hqontario.ca/public-reporting/patient-safety).
20. Canadian classification of health interventions. Vol. 3: tabular list. Ottawa: Canadian Institute of Health Information, 2003.
21. The Johns Hopkins ACG Case-Mix System, version 8.2. Baltimore: Johns Hopkins Bloomberg School of Public Health, 2008.
22. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988;44:1049-60. [Erratum, Biometrics 1989;45:347.]
23. Surgical Safety Checklist "how-to" implementation guide. Toronto: Ontario Hospital Association (https://www.oha.com/CurrentIssues/keyinitiatives/PatientSafety/Documents/SSC_Toolkit_PartB_Feb04.pdf).
24. Levy SM, Senter CE, Hawkins RB, et al. Implementing a surgical checklist: more than checking a box. Surgery 2012;152:331-6.
25. World Alliance for Patient Safety. Implementation manual: Surgical Safety Checklist. Geneva: World Health Organization (http://www.who.int/patientsafety/safesurgery/tools_resources/SSSL_Manual_finalJun08.pdf).
26. North Carolina Quality Center. WHO Surgical Safety Checklist. Cary, NC: NCQC (http://www.ncqualitycenter.org/?s=WHO+surgical+safety+checklist).
27. Safe Surgery 2015: South Carolina (http://www.safesurgery2015.org/safe-surgery-2015-south-carolina.html).
28. Oregon Association of Hospitals and Health Systems. Sprint for surgical safety (http://www.oahhs.org/quality/initiatives/sprint-to-surgical-safety.html).
29. Auerbach AD, Landefeld CS, Shojania KG. The tension between needing to improve care and knowing how to do it. N Engl J Med 2007;357:608-13.
30. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 2010;14:1-193.
31. Tu JV, Wu K. The improving outcomes of coronary artery bypass graft surgery in Ontario, 1981 to 1995. CMAJ 1998;159:221-7.
32. Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. JAMA 2010;303:1259-65.
33. Hu JC, Gu X, Lipsitz SR, et al. Comparative effectiveness of minimally invasive vs open radical prostatectomy. JAMA 2009;302:1557-64.
34. Schermerhorn ML, O'Malley AJ, Jhaveri A, Cotterill P, Pomposelli F, Landon BE. Endovascular vs. open repair of abdominal aortic aneurysms in the Medicare population. N Engl J Med 2008;358:464-74.
35. Cima RR, Lackore KA, Nehring SA, et al. How best to measure surgical quality? Comparison of the Agency for Healthcare Research and Quality Patient Safety Indicators (AHRQ-PSI) and the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) postoperative adverse events at a single institution. Surgery 2011;150:943-9.
36. Best WR, Khuri SF, Phelan M, et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. J Am Coll Surg 2002;194:257-66.
37. Kaafarani HM, Rosen AK. Using administrative data to identify surgical adverse events: an introduction to the Patient Safety Indicators. Am J Surg 2009;198:Suppl:S63-S68.

Base deficit as an early marker of coagulopathy in trauma

S Cheddie, MMed (Surg), FCS (SA); **D J J Muckart,** FRCS, MMSc; **T C Hardcastle**, MMed, FCS (SA)

Department of General Surgery, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa

Corresponding author: *D J J Muckart (davidmuc@ialch.co.za)*

Background. The acute coagulopathy of trauma is associated with hypoperfusion, metabolic acidosis and an increased mortality rate. Biochemical markers of hypoperfusion, namely base deficit (BD) and lactate, are commonly used to assess the degree of hypoperfusion. Early identification of hypoperfusion and acidosis using BD and lactate may help predict the development of coagulopathy in trauma patients and direct therapy.

Objectives. To identify whether a correlation exists between BD, lactate, injury severity, early-onset coagulopathy and mortality.

Methods. A retrospective chart analysis was undertaken of patients transferred directly from scene to the level I trauma unit at Inkosi Albert Luthuli Central Hospital, Durban, South Africa, from 2007 to 2008. Patients with evidence of hypoperfusion were selected. Hypoperfusion was defined as a base deficit >-2 and coagulopathy as an International Normalized Ratio (INR) of >1.2. BD, lactate, chloride, temperature, Injury Severity Score (ISS), INR and mortality were recorded in this cohort. Student's *t*-test and Fisher's exact test were used for continuous and categorical variables, respectively. Correlation curves were used to determine the degree of association between the variables BD, lactate and ISS with respect to the INR. A *p*-value of <0.05 was considered statistically significant.

Results. Of the 28 patients, males (n=18) accounted for 64.3% of admissions. The mean age was 31 years (range 1 - 75 years, median 30 years). The mechanism of injury was penetrating trauma in 5 cases (17.9%) and blunt trauma in 23 (82.1%). The median ISS was 24 (range 4 - 59). In 16 patients (57.1%) the INR was within normal limits, but in 12 (42.9%) it was over 1.2. There was a significant correlation between BD, ISS and INR (r=0.393; p=0.019 and r=0.565, respectively; p<0.001). Lactate showed a weak and non-significant association with the INR (r=0.232; p=0.18). There were a total of 12 deaths (42.8%) in this cohort of patients with biochemical evidence of hypoperfusion. There was a significant increase in mortality in patients with evidence of hypoperfusion and an elevated INR (75.0% v. 18.7%; p=0.006).

Conclusion. BD but not lactate correlates with the development of the coagulopathy of trauma. The ISS showed a significant correlation with coagulation disturbances, and the combination of hypoperfusion and coagulopathy was associated with a significant increase in mortality.

S Afr J Surg 2013;51(3):88-90. DOI:10.7196/SAJS.1665

The acute coagulopathy of trauma is a complex pathophysiological state that may be initiated by a number of factors. As an integral part of the 'deadly triad', it may be perpetuated by concomitant acidosis and hypothermia;[1] excessive pre-hospital fluid resuscitation may lead to haemodilution of clotting factors and worsening of the coagulopathy; and exposure of tissue factor when endothelium or brain matter is denuded results in release of thromboplastins.[2] Hypoperfusion following major injury is associated with an increased expression of thrombomodulin, which binds thrombin, resulting in increased activation of protein C that subsequently inhibits plasminogen activator inhibitor, leading to increased fibrinolysis.[3]

The coagulopathy of trauma is associated with an increased mortality rate,[4] and the key to improving survival lies in identifying and correcting early potentially reversible risk factors. Although traditionally hypoperfusion has been diagnosed and classified on the basis of clinical vital signs, occult hypoperfusion may occur in the presence of normal haemodynamic parameters. The measurement of lactate, the product of anaerobic metabolism, and the surrogate base deficit (BD) are well-recognised markers of cellular hypoperfusion and the severity of shock,[5] although BD may also be elevated under aerobic conditions such as acute kidney injury or hyperchloraemia, both of which commonly occur after major trauma as a result of renal hypoperfusion and excessive saline-containing resuscitation fluids. Both lactate and BD have been extensively investigated in trauma patients as markers of injury severity, end-points of resuscitation and predictors of outcome.

Although indicating the severity of hypoperfusion, measurement of lactate alone may not portray the extent of the metabolic acidosis. BD quantifies the extent of both anaerobic and aerobic acidoses and may be a better indicator of the risk of a coagulation disturbance. This study was undertaken to identify the correlation between BD, lactate and injury severity with the development of early-onset coagulopathy and mortality.

Methods

The study was approved by the Biomedical Research Ethics Committee (BE207/09) of the University of KwaZulu-Natal, Durban, South Africa. A retrospective chart analysis of 75 trauma patients admitted directly from the scene of injury to the level I trauma unit at Inkosi Albert Luthuli Central Hospital, Durban, from 2007 to 2008 was performed and data pertaining to patients with evidence of hypoperfusion were extracted. Hypoperfusion was defined as a BD >-2 (for the purpose of this study the symbol 'greater than' describes a more negative BD) and coagulopathy as an International Normalized Ratio (INR) of >1.2. All blood results were those taken on admission to the resuscitation room via an arterial sample. The BD was obtained from arterial blood gas analysis. Arterial lactate, chloride ion concentration and INR results were extracted from the Medicom electronic hospital laboratory database. Tympanic membrane temperature was recorded on admission to the resuscitation room. The Injury Severity Score (ISS) was computed after all injuries had been identified using the Abbreviated Injury Scale (AIS) 90 reference book. Death or survival was ascertained from the electronic hospital discharge notes. Complete data were available for 28 patients and were included in the analysis.

Student's *t*-test and Fisher's exact test were used for continuous and categorical variables, respectively. Correlation curves were used to determine the degree of association between the variables BD, lactate and ISS with respect to the INR. A *p*-value of <0.05 was considered statistically significant.

Results

Of the 28 patients, 18 were males (64.3%) and 10 females (35.7%). The mean age was 31 years (range 1 - 75 years, median 30 years). The mechanism of injury was penetrating trauma in 5 cases (17.9%) and blunt trauma in 23 (82.1%). The median ISS was 24 (range 4 - 59). In 16 patients (57.1%) the INR was within normal limits, but in 12 (42.9%) it was >1.2. The differences in mean BD, lactate, INR, ISS, systolic blood pressure and temperature between these two groups are shown in Table 1.

Lactate and BD showed a strong and significant correlation (r=0.739; p<0.001). There was a moderate and significant correlation between BD and INR (r=0.393; p=0.019), and the ISS demonstrated a stronger significant association (r=0.565; p<0.001). Lactate showed a weak and non-significant association with the INR (r=0.232; p=0.18). Chloride was above the upper limit of normal in 17 patients (60.7%), although there was no significant difference between the coagulopathic and non-coagulopathic patients.

There were a total of 12 deaths (42.8%) in this cohort of patients with biochemical evidence of hypoperfusion. In the group with coagulopathy 9/12 (75.0%) died, compared with only 3/16 (18.7%) of those with a normal INR (p=0.006).

Discussion

Metabolic acidosis has been incriminated as one of the instigators of coagulation disturbances following injury. The BD is defined as the amount of milli-equivalents of base required to titrate a litre of blood to a pH of 7.4 at an arterial partial pressure of carbon dioxide ($PaCO_2$) of 40 mmHg.[6] This measurement is unaffected by acute changes in $PaCO_2$, is universally elevated in the presence of all pathological changes that induce a metabolic acidosis, and as such is a more reliable marker of the severity of the underlying metabolic state than pH, which will be altered by respiratory compensation. In the trauma setting, increased lactate production arises from anaerobic metabolism as a consequence of hypoperfusion, most commonly due to haemorrhage, but will not be affected by a metabolic acidosis that arises under aerobic conditions. Despite the fact that the BD not only reflects the degree of lactate accumulation but also provides a more composite view of a patient's metabolic status,[6,7] lactate clearance has been more popular than BD as a predictor of mortality in the trauma population.[8] Davis[9] showed a strong correlation between BD and lactate in a porcine haemorrhagic shock model and concluded that the BD was as accurate a marker of tissue hypoperfusion as lactate.The admission BD has also been shown to correlate well with the ISS, development of multiple organ failure, length of intensive care unit and hospital stay, need for blood transfusions and mortality,[10] and a BD of ≥-6 has been shown to be a marker of injury severity and mortality in the trauma population, irrespective of age.[3,10,11]

Although hypoperfusion and the subsequent lactic acidosis play a key role in the pathogenesis of the early coagulopathy of trauma, the addition of another mechanism for metabolic acidosis will accentuate the risk. Our results show a moderate but significant correlation between an elevated BD and coagulopathy as measured by the INR, but we could not demonstrate a similar association with lactate despite the BD and lactate showing a strong correlation. As mentioned above, the reason may lie in the fact that lactate only reflects an acidosis arising from anaerobic metabolism, whereas the BD is a global marker of both anaerobic and aerobic acidoses. In addition to (and as a consequence of) hypoperfusion, acute

Table 1. Comparison of coagulopathic v. non-coagulopathic patients

	Coagulopathy (mean±SD) (*n*=12)	No coagulopathy (mean±SD) (*n*=16)	*p*-value
Lactate (mmol/l)	7.04±7.18	3.69±2.98	0.10
BD (mmol/l)	-9.46±4.47	-5.82±2.50	0.011
INR	1.58±0.17	1.08±0.07	<0.001
Temperature (°C)	35.9±1.1	36.3±1.2	0.34
Systolic BP (mmHg)	110.4±41.8	122.6±27.2	0.36
ISS	32.2±16.2	17.8±9.8	0.007

SD = standard deviation; BD = base deficit; INR = International Normalized Ratio; BP = blood pressure; ISS = Injury Severity Score.

kidney injury may contribute to the acidosis of injury over and above that from anaerobic metabolism. Hyperchloraemia is a common consequence of kidney injury, and although previously thought innocuous has a detrimental effect on outcome when combined with another acidosis.[12] Furthermore, hyperchloraemia may not only contribute to an acidotic state during renal insufficiency, but aggravate renal damage further by inducing renal vasoconstriction.[13,14] Given the incidence of hyperchloraemia in our patient population, which in addition to hypoperfusion may contribute to metabolic acidosis, acute kidney injury and coagulopathy, the BD may be a more useful predictor of the risk of coagulation disturbances and an indicator for the early use of plasma rather than clear fluids for resuscitation.

The ISS showed the strongest association with an elevated INR. Although hypoperfusion is undoubtedly a contributing factor in the coagulopathy of trauma, it is not a prerequisite. Tissue factor exposure has been incriminated as a major trigger mechanism,[2,15] and it may be assumed that a higher ISS would be associated with more severe tissue damage, an increase in coagulation disturbances and a higher mortality rate. The presence of coagulopathy at admission is associated with increased injury severity, length of hospital stay and number of organ failures and mortality,[16] and we have shown a statistically significant increase in mortality in our cohort of patients with evidence of hypoperfusion and who were coagulopathic on admission.

In the acute resuscitation phase, acquiring an INR result may take up to an hour. In the emergency setting, thromboelastometry has been shown to be a quick, useful method in identifying patients with coagulopathy and guiding transfusion requirements. A strong correlation has been shown between thromboelastometry and derangements in conventional clotting parameters, results are available within 10 minutes, and this simple device has now gained acceptance as standard of care in the severely injured.[17]

It is imperative to identify patients with acute coagulopathy early as this may influence outcome and resuscitation efforts. The BD can be obtained within minutes using a conventional blood gas analyser, which is available in most emergency departments, and has a significant association with acute coagulation disturbances. This finding could have far-reaching clinical applications. In addressing hypoperfusion in the resuscitation bay, the conventional use of crystalloids or colloids could be replaced by the immediate use of freeze-dried plasma that would lead to early replenishment of clotting factors in addition to sustained volume expansion. The recent war in Iraq has given rise to similar opinions, suggesting that fresh-frozen plasma be used as the resuscitation fluid of choice in the acute combat setting to decrease additional bleeding and replenish clotting factor deficiencies.[18]

There are several limitations to our study. The sample may be too small in size and not fully representative of the trauma population. There are major discrepancies in the literature with regard to defining coagulopathy, and the INR value of >1.2 that we used may be too sensitive and overestimate the presence of coagulopathy in the study group. In a review by Brohi *et al.*, four major studies used four different indices to define coagulopathy.[15] An estimate of pre-hospital fluid administration was lacking from our database and was not included in our assessment. Even considering these shortfalls, there is a growing body of evidence indicating that in order to improve survival in the seriously injured we need to address the possibility of a coagulation disturbance much earlier than was previously thought. The simple combination of BD and an estimate of the ISS will help to identify those at risk.

Conclusion

Our findings suggest that a significant correlation exists between BD, ISS and INR and that BD is superior to lactate in predicting coagulation disturbances. The combination of hypoperfusion and a coagulopathy at admission is associated with a significant increase in mortality.

REFERENCES

1. Tieu BH, Holcomb JB, Schreiber MA. Coagulopathy: Its pathophysiology and treatment in the injured patient. World J Surg 2007;31(5):1055-1064. [http://dx.doi.org/10j007/s00268-006-0653-9]
2. Ho AM, Karmakar MK, Dion PW. Are we giving enough coagulation factors during major trauma resuscitation? Am J Surg 2005;190(3):479-484. [http://dx.doi.org/10.1016/j.amjsurg.2005.03.034]
3. Brohi K, Cohen MJ, Ganter MT, et al. Acute coagulopathy of trauma: Hypoperfusion induces systemic anticoagulation and hyperfibrinolysis. J Trauma 2008;64(5):1211-1217. [http://dx.doi.org/10.1097/TA0b013e318169cd3c]
4. Brohi K, Singh J, Heron M, Coats T. Acute traumatic coagulopathy. J Trauma 2003;54(6):1127-1130. [http://dx.doi.org/10.1097/01.TA0000069184.06]
5. Paladino L, Sinert R, Wallace D, Anderson T, Yadav K, Zehtabchi S. The utility of base deficit and arterial lactate in differentiating major from minor injury in trauma patients with normal vital signs. Resuscitation 2008;77(1):363-368. [http://dx.doi.org/10.1016/j.resuscitation.2008.01.022]
6. Juern J, Khatri V, Weigelt J. Base excess: A review. J Trauma Acute Care Surg 2012;73(1):27-31. [http://dx.doi.org/10.1097/TA.0b013e318256999d]
7. FitzSullivan E, Salim A, Demetriades D, Arseno J, Martin MJ. Serum bicarbonate may replace the arterial base deficit in the trauma intensive care unit. Am J Surg 2005;190(6):961-967. [http://dx.doi.org/10.1016/j.amsurg.2005.08.024]
8. Abramson D, Scalea TM, Hitchcock R, Trooskin SZ, Henry SM, Greenspan J. Lactate clearance and survival following injury. J Trauma 1993;35(4):584-589. [http://dx.doi.org/10.1097/00005373-199310000-00014]
9. Davis JW. The relationship of base deficit to lactate in porcine hemorrhagic shock and resuscitation. J Trauma 1994;36(2):168-172. [http://dx.doi.org/10.1097/00005373-199402000-00002]
10. Davis JW, Parks SN, Kaups K, Gladen H, O'Donnel-Nichol SRN. Admission base deficit predicts transfusion requirements and risk of complications. J Trauma 1996;41(5):769-774. [http://dx.doi.org/10.1097/00005373-199611000-00001]
11. Rutherford EJ, Morris JA, Reed GW, Hall KS. Base deficit stratifies mortality and determines therapy. J Trauma 1992;33(3):417-423. [http://dx.doi.org/10.1097/00005373-199209000-00014]
12. Shaw AD, Bagshaw SM, Goldstein SL, et al. Major complications, mortality, and resource utilization after open abdominal surgery. Ann Surg 2012;255(5):821-829. [http://dx.doi.org/10.1097/SLA.0b013e31825074f5]
13. Yunous NM, Bellomo R, Story D, Kellum J. Bench-to-bedside review: Chloride in critical illness. Crit Care 2010;14(4):226-235. [http://dx.doi.org/10.1186/cc9052]
14. Yunous NM, Bellomo R, Hegarty C, et al. Association between a chloride-liberal vs chloride-restrictive intravenous fluid administration strategy and kidney injury in critically ill adults. JAMA 2012;308(15):1556-1572. [http://dx.doi.org/10.1001/jama.2012.13356]
15. Brohi K, Cohen MJ, Davenport RA. Acute coagulopathy of trauma: Mechanism, identification and effect. Current Opinion in Critical Care 2007;13(6):680-685. [http://dx.doi.org/10.1097/MCC.0b013e3282f1e78f]
16. Maegele M, Lefering R, Yucel N, et al. Early coagulopathy in multiple injury: An analysis from the German Trauma Registry on 8724 patients. Injury 2007;38(3):298-304. [http://dx.doi.org/10.1016/j.injury.2006.10.003]
17. Rugeri L, Levrat A, David JS, et al. Diagnosis of early coagulation abnormalities in trauma patients by rotation thrombelastography. J Thromb Haemost 2007;5(2):289-295. [http://dx.doi.org/10.1111/j.1538-7836.2007.02319.x]
18. Sebesta MAJ. Special lessons learned from Iraq. Surg Clin North Am 2006;86(3):711-772. [http://dx.doi.org/10.1016/j.suc.2006.03.002]