

**Where do the (big) data  
come from?**

**... and why it matters**

**3. Comparing treatments:  
key concepts**

Rod Little

Department of Biostatistics



# Two studies with very different conclusions

- Two articles on treatment of advanced cancer using Vitamin C yield conflicting conclusions:
- Cameron, E. and Pauling, L. (1976). Supplemental ascorbate in the supportive treatment of cancer: prolongation of survival times in terminal human cancer Proc. Natl. Acad. Sci. USA Vol. 73, No. 10, pp. 3685-3689, 1976.
- Creagan, E. et al (1979). Failure of High Dose Vitamin C (ascorbic acid) therapy to benefit patients with advanced cancer. New. Eng. J. Med. 301: 687-690, 1979.

# Questions

- Cameron & Pauling: large effect of Vit C
- Creagan et al.: no effect of Vit C
- Why do these studies give such different results, and which should we believe?
- The U.S. Food and Drug Administration (FDA) (and international equivalents) decide when treatments should be approved for widespread use
  - a big responsibility not to sanction treatments that are harmful, or stand in the way of treatments that are beneficial
- Major role of study design

# Evidence-based medicine

- The idea that choices between different treatments or behaviors should be based on empirical evidence, rather than opinions of “experts”
- Plausible theories can often be provided for effectiveness of many treatments – see e.g. the Cameron and Pauling arguments for Vitamin C as a treatment of cancer
- While scientific plausibility is important, empirical evidence is key, since “plausible” does not necessarily mean “right”

# Data, Data, everywhere!

- We use data to answer public health questions
  - Effectiveness of treatments for cancer
  - Effectiveness of COVID 19 treatments and vaccines
  - Relationships between pollutants and health outcomes
- How strong is the evidence?
  - Many studies have conflicting conclusions
  - Design: How were the data collected? What are the strengths and weaknesses of various studies?
  - GIGO (Garbage In, Garbage Out). Clever statistical analysis can't rescue an inherently flawed study.
- Statistical analysis
  - Distinguish real from chance differences.
  - But the design is crucial for assessing whether significant differences are “causal” – caused by the treatment rather than other factors (confounders)?

# Badly designed studies can do serious harm!

- Vaccines and autism
- “In recent years the antivaccine movement has focused on the claim that vaccines are linked to neurological injury, and specifically to the neurological disorder autism, now referred to as autism spectrum disorder (ASD). However the scientific evidence overwhelmingly shows no correlation between vaccines in general, the MMR vaccine specifically, or thimerosal (a mercury-based preservative) in vaccines with ASD or other neurodevelopmental disorders.”  
<https://www.sciencebasedmedicine.org/reference/vaccines-and-autism/>
- See our skit: “Just the Vax Ma’am!”
- <https://m.youtube.com/watch?feature=youtu.be&v=ag5bjEpxnaM>
- Vaccine hesitancy has led to measles outbreaks, and is lengthening the COVID 19 pandemic in the US

## Early report

## Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

## Summary

**Background** We investigated a consecutive series of children with chronic enterocolitis and regressive developmental disorder.

**Methods** 12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent gastroenterological, neurological, and developmental assessment and review of developmental records. Ileocolonoscopy and biopsy sampling, magnetic-resonance imaging (MRI), electroencephalography (EEG), and lumbar puncture were done under sedation. Barium follow-through radiography was done where possible. Biochemical, haematological, and immunological profiles were examined.

**Findings** Onset of behavioural symptoms was associated by the parents, with measles, mumps, and rubella vaccination in eight of the 12 children, with measles infection in one child, and otitis media in another. All 12 children had intestinal abnormalities (ranging from lymphoid nodular hyperplasia to granuloid ulceration). Histology showed patchy chronic inflammation in seven, but no granulomas. Behavioural disorders included autism (nine), disintegrative psychosis (one), and possible postviral or vaccinal encephalitis (two). There were no focal neurological abnormalities and MRI and EEG tests were normal. Abnormal laboratory results were significantly raised urinary methylmalonic acid compared with age-matched controls ( $p=0.03$ ), low haemoglobin in four children, and low serum IgA in four children.

**Interpretation** The idiopathic associated gastrointestinal disease and developmental regression in a group of previously normal children, which was generally associated in time with possible environmental triggers.

*Lancet* 1998; **351**: 637–41

See Comment *by page*

**Inflammatory Bowel Disease Study Group, University Departments of Medicine and Histopathology** (A J Wakefield *FRCS*, A Anthony *MR*, J Linnell *FRCD*, A P Dhillon *MRCPsib*, S E Davies *MRCPsib*) **and the University Departments of Paediatric Gastroenterology** (S H Murch *MR*, D M Casson *MRCP*, M Malik *MRCP*, M A Thomson *FRCP*, J A Walker-Smith *FRCP*), **Child and Adolescent Psychiatry** (M Berelowitz *FRCPsib*), **Neurology** (P Harvey *FRCP*), **and Radiology** (A Valentine *FRCD*), **Royal Free Hospital and School of Medicine, London NW3 2QG, UK**

Correspondence to: Dr A J Wakefield

The source of the vaccine-autism link is this (very poorly designed) study

## Introduction

We saw several children who, after a period of apparent normality, lost acquired skills, including communication. They all had gastrointestinal symptoms, including abdominal pain, diarrhoea, and bloating and, in some cases, food intolerance. We describe the clinical findings, and gastrointestinal features, of these children.

## Patients and methods

12 children, consecutively referred to the department of paediatric gastroenterology with a history of a pervasive developmental disorder with loss of acquired skills and intestinal symptoms (abdominal pain, bloating and food intolerance), were investigated. All children were admitted to the ward for a week, accompanied by their parents.

## Clinical investigations

We took histories including details of immunisations and exposure to infectious diseases, and assessed the children. In 11 cases the history was obtained by the senior clinician (JW-S). Neurological and psychiatric assessments were done by consultant staff (PH, MB) with HMS-4 criteria.<sup>1</sup> Developmental records included a review of prospective developmental records from parents, health visitors, and general practitioners. Four children did not undergo psychiatric assessment in hospital; all had been assessed professionally elsewhere, so these assessments were used as the basis for their behavioural diagnosis.

After bowel preparation, ileocolonoscopy was performed by SHM or MAT under sedation with midazolam and pethidine. Paired frozen and formalin-fixed mucosal biopsy samples were taken from the terminal ileum; ascending, transverse, descending, and sigmoid colons, and from the rectum. The procedure was recorded by video or still images, and were compared with images of the previous seven consecutive paediatric colonoscopies (four normal colonoscopies and three on children with ulcerative colitis), in which the physician reported normal appearances in the terminal ileum. Barium follow-through radiography was possible in some cases.

Also under sedation, cerebral magnetic-resonance imaging (MRI), electroencephalography (EEG) including visual, brain stem auditory, and sensory evoked potentials (where compliance made these possible), and lumbar puncture were done.

## Laboratory investigations

Thyroid function, serum long-chain fatty acids, and cerebrospinal-fluid lactate were measured to exclude known causes of childhood neurodegenerative disease. Urinary methylmalonic acid was measured in random urine samples from eight of the 12 children and 14 age-matched and sex-matched normal controls, by a modification of a technique described previously.<sup>2</sup> Chromatograms were scanned digitally on computer, to analyse the methylmalonic-acid zones from cases and controls. Urinary methylmalonic-acid concentrations in patients and controls were compared by a two-sample *t* test. Urinary creatinine was estimated by routine spectrophotometric assay.

Children were screened for antiendomysial antibodies and boys were screened for fragile-X if this had not been done

# Key concepts

We focus on the following key concepts:

1. Defining a causal effect – the Rubin/Neyman causal model
2. Confounding and internal validity
3. Effect-modification and external validity
4. Alternative study designs and their strengths and weaknesses – in particular, the role of randomization in the assignment of treatments



# Goals of Research Design

- Internal validity: are the estimated effects of the treatments valid for the individuals in the study?
  - A crucial component — avoiding bias of all kinds
- External validity/Generalizability: are the estimated effects valid for the target population of to which the treatments are to be applied
  - internal validity is a prerequisite
  - Individuals in a study are usually volunteers, not randomly sampled from the target population -- does that matter?
  - There's a tendency to leap to inference far beyond the targeted population.

# When is a treatment effect causal?

- How do we know the improvement is caused by the treatment and not something else?
- This gets to a central question: how do we define a causal effect? Phenomena have multiple causes, often hard to disentangle...
- E.g. what “causes” mass shootings
  - Ready access to guns, lack of gun training, mental health of shooters, etc. etc.

# Defining causal effects

- Association is not causation: We are interested in causal effects of treatments/etiologic factors.
  - How do we define a “causal effect”?
- “Rubin Causal Model” – causal effect of treatment for subject is difference in outcome under active treatment and under control.
- Estimation of causal effects is basically a missing data problem: We only get to see the outcome from one treatment, the treatment actually received!
- How the treatments are assigned is a crucial issue – randomization plays a key role in avoiding bias

# Numerical example

$Y(j)$  = depression score given treatment  $j$   
 (high = more depressed)

Subject	Y(A)	Y(B)	Y(A)-Y(B)
1	[1]	6    6	[-5]
2	[3]	12   12	[-9]
3	9    9	[10]	[-1]
4	11   11	[12]	[-1]
Mean	10* [6]	9* [10]	1* [-4]

- Assignment mechanism is confounded:  
 Sicker (more depressed) subjects got treatment A!

need for a comparison group

# Confounding

- $X_2$  is a confounding factor for effect of treatment  $X_1$  on  $Y$  if it is not an outcome of treatment, its distribution differs between treatments, and it affects the outcome
  - Confounding is an important issue for *internal validity*: whether a treatment effect is causal for the individuals in a study.
  - In numerical example, baseline depression is a confounding variable

# Assignment mechanism

$$T = \begin{cases} A, & \text{if assigned to treatment A} \\ B, & \text{if assigned to treatment B} \end{cases}$$

$Y(A)$  = Outcome if assigned A

$Y(B)$  = Outcome if assigned B

- Assignment mechanism is called *unconfounded* if

$$T \perp [Y(A), Y(B)], \quad \perp = \text{independent}$$

Otherwise assignment mechanism is confounded

- Average causal effects can be estimated as difference in observed means if *assignment* mechanism is *unconfounded*

$$E[Y | T = j] = E[Y(j)]$$

# Alternative Designs

- Suppose we have a new treatment, and we want to assess its effectiveness
- (Or: we are interested in whether an environmental factor is causally related to disease)
- Consider alternative designs:
  - “Snake Oil Salesman” (SOS)
  - Other observational designs
  - Randomized Clinical Trial (RCT)

# The SOS Design

- Give someone the treatment and see if they get better
- Seems logical
- I call this the “Snake-Oil Salesman” (SOS) design
- Much seen in “before and after” commercials on TV



# Need for comparison group

- Why not simply assign the new treatment to everyone in study and see if they improve?
  - Do not observe outcome under “no treatment”
  - Implicitly makes dubious assumption of no change under no treatment
  - Better designs have a comparison group.

# Three more problems with SOS

- Selection bias: even if the treatment does nothing, if the outcome is variable, we can cherry-pick the cases where the outcome improved
  - E.g. weight loss on a diet – after the diet starts, some people lose weight, some gain weight, some don't change much. Select the ones that lose weight
  - Investment managers etc.: the ones that flog books on TV are the ones that made money, but it could be they were not smart, just lucky
  - History is written by the winners...
  - see “Fooled by Randomness” by Nassim Taleb

# Three more problems with SOS

- Regression to the mean: if the outcome is change in a measure (e.g. depression) and that measure fluctuates naturally, then people who start high on the measure will tend finish lower, and people who start low on the measure will tend to finish higher, without any treatment
- E.g. baseball: after 20 at bats, some players are batting .100 (2 hits) and some are batting .600 (12 hits)
- After 200 bats, those batting .100 will in all likelihood end up higher, and those batting .600 will end up lower
- If we select individuals batting .100, and give them a magic “batting snake oil” they’ll surely improve, even though the improvement has nothing to do with the oil

# Three more problems with SOS

- Placebo effect: even in the absence of any active ingredient, people report an improvement.
- If a treatment involves an investment, we want to believe the investment has been worthwhile – not throwing time or money down the drain – hence believe the treatment has worked
- Particularly a problem with subjective responses, like pain scores; objective measures are less vulnerable

# Case Reports and Case Series

- Similar in nature to the SOS design are reports of unusual medical occurrences or associations:
  - Led to early identification of the AIDS epidemic
  - Useful in identifying unusual clusters of disease
- Hypothesis generating
- Anecdotal; not valid statistical evidence
- Sometimes it's real:
  - Vinyl chloride and liver disease
- Sometimes it's not:
  - Breast implants and scleroderma

# Example: Disease clusters

- Newspaper reports that 4 out of 8 pregnant female secretaries in a large office with extended exposure to electromagnetic radiation from computer monitors had spontaneous abortions!
- Causality or coincidence?
- Worrying, but newspaper could be reporting a chance event in the tail of the distribution — what about the thousands of offices where this surprising number of abortions did not occur?
- Need prospective clinical study to avoid selection bias

# Cross-sectional Surveys

- Exposure and disease status are assessed at a single survey. For example:
  - Assessing fluoride history and number of dental cavities at a single visit
  - National health and nutrition examination survey (NHANES)
- Such studies often find associations between disease and exposure.
- But, is the association truly causation?
  - E.g., did the exposure precede the disease?
  - E.g., does sedentary lifestyle cause CHD, or do people with developing CHD feel too ill to exercise?

# Prospective Observational Studies

- The problems with the SOS design suggest that we need a comparator – a placebo, or an existing treatment
  - Some individuals are assigned the new treatment, and some are assigned the comparator treatment.
- Compare two groups with respect to an appropriate outcome, e.g. five year survival rates, and see which group does better
- BUT: If assignment to treatment/comparator is not random, there may be confounding factors.



# What's an observational study?

- Assignment of the treatment or etiological factor is natural and not under the control of the investigator
  - Environmental factors are not randomly assigned
  - Smoking is choice of the study participant
  - Treatments in clinical data bases are assigned by clinicians, not controlled by the researcher
  - Review of historical case records

# Confounding in Observational Studies

- Inference from every observational study depends on eliminating bias and adjusting for all confounding factors.
  - Confounding factors: age, gender, income, disease severity, etc. may be correlated with the treatment assignment and predict the outcome
- Analysis methods can (multiple and logistic regression, propensity adjustment) can adjust for observed confounders.
- But unobserved confounders remain a problem

# Example: learning health systems

- An administrative health system captures data for 200 patients with a rare disorder – 100 are taking Drug A and 100 drug B. 70 people taking Drug A are “cured” and 30 people taking Drug B are “cured”
- The naïve conclusion is that Drug A is more effective. [Note: this difference too large to be attributable to chance]
- But we can’t conclude that Drug A is better – maybe something other than the effect of the drug – a confounding factor -- explains the difference...
- For valid inference, need to record and adjust for potential confounders in the analysis

# Crossover designs

- An approximation to observing outcome under both treatments is achieved in crossover designs
  - Individuals receive both treatments A and B, and outcome is recorded for both.
  - Need to guard against spillover effects by suitable “washout period” between treatments
  - Good when feasible, but only possible for short-term, treatment of chronic conditions
  - Randomizing the order of treatments (A then B or B then A) is a good idea to reduce “order effects”.
  - Still short of ideal — conditions under which treatments are given are still not identical.

# Case-Control Studies

- Cases with disease are identified; controls are selected from the same population that gave rise to the cases.
- The proportions exposed among cases and controls are compared.
  - E.g., compare the proportion of smokers among lung cancer patients and non-cancer controls.
- An efficient design for rare diseases
  - In a simple random sample, lung cancer cases would be quite rare, so a huge sample size would be needed to make the same comparison.
- Assignment not at random, may be confounded

# Selecting Controls

- The hardest and most important design issue. Controls are selected from the population that gave rise to the cases.
- Hospital controls: convenient, cheap
  - Use other patients, without the target disease.
  - Because they are ill, they have been shown to be different from the general population (e.g., more likely to smoke and be heavy drinkers).
- Population controls: the gold standard
  - RDD or canvassing households
- Friend / neighbor / relative controls

# Potential Bias in Exposure Ascertainment

- Information from record reviews
  - May have missing or incorrect information
  - Case info may be more completely documented.
- Patient interviews
  - Different response rates in cases and controls
    - Cases may be more willing to participate
  - Recall bias
    - Differential reporting of exposure in cases and controls
    - For long-ago exposures, memory helpers (e.g., concurrent residential history) may be helpful.
    - Make sure the exposure pre-dated the disease

# Randomized Clinical Trials

- Random assignment of subjects to treatments yields an unconfounded assignment mechanism
  - Facilitates causal inference.
  - Eliminates selection bias from choosing the “best” patients for the preferred treatment



# RCT's vs. Observational Studies

- Randomized clinical trials
  - Assignment is random, hence unconfounded
- Observational studies (e.g., registries)
  - Assignment of treatment is uncontrolled, potentially confounded
  - Easier to conduct
  - Good for hypothesis generation
  - Necessary when randomization cannot be performed

# Randomized assignment

- All participants are treated the same, except for the treatment assigned
- Unconfounded assignment mechanism, eliminates observed and unobserved confounding factors
  - including the investigator's conflict of interest in favor of new treatment
  - Blinding to treatment, if feasible, removes potential bias in whether or not participants are included

# Blinding / Masking

- Single-blind: The patient does not know which treatment s/he is receiving.
- Double blind: Both patient and investigator do not know the treatment assignment.
- Triple blind: The person analyzing the data is also masked to the treatment assignment.
- The evaluator may be a different person, and blinding of this person is crucial.

# Blinded Studies (cont'd)

- Blinding removes or equalizes biases due to patients' desire to please and investigator enthusiasm.
- Logistics:
  - Blinded studies of drugs are simple because placebo pills can usually be made.
  - Blinded studies of surgery vs. medical management are hard, sometimes not possible. (But see later).

# Levels of evidence

- Several groups have attempted to provide “levels of evidence” for medical study designs. See for example [https://en.wikipedia.org/wiki/Levels\\_of\\_evidence](https://en.wikipedia.org/wiki/Levels_of_evidence)
- <http://www.ahfsdruginformation.com/levels-of-evidence-rating-system/>

Double-blind RCT’s are generally considered the gold standard, when feasible

# Article Critique 1

- The following outline serves as a framework for evaluating articles in the public health literature.
- 1. General
  - Experiment or survey?
  - What are the authors seeking to demonstrate? Are they consistent?
- 2. Sample Selection
  - To what population (are/can) their results to be generalized?
  - Biases introduced by selection of cases? (nonresponse, excluded cases)
  - Sample large enough? Sufficient statistical power to detect differences of substantive interest?

# Article Critique 2

- 3. Treatment Allocation
  - Sufficient documentation ?
  - What evidence is there that treatment arms are equal except for treatments applied:
    - Randomized allocation of treatments?
    - Stratification?
    - Treatment groups compared on observed factors?
    - Might unobserved factors explain the difference in outcomes?
    - Blinding (masking) (of subjects, treatment administrators, investigators )? Possible? Done?
    - Placebo effect?

# Article Critique 3

- 4. Outcome Measures
  - Appropriate?
  - Clearly defined and reproducible?
  - Affect all treatment arms equally?
- 5. Analysis of Results
  - Adequate presentation of data?
  - Appropriate statistical analyses?
  - Arithmetic errors? Do the results look right?
  - Appropriate inferences from the analysis?
  - Balanced conclusions?
- 6. Constructive Criticism



# Example: Vitamin C and Cancer

- Two articles on treatment of advanced cancer using Vitamin C yield conflicting conclusions:
- Cameron, E. and Pauling, L. (1976). Supplemental ascorbate in the supportive treatment of cancer: prolongation of survival times in terminal human cancer Proc. Natl. Acad. Sci. USA Vol. 73, No. 10, pp. 3685-3689, 1976.
- Creagan, E. et al (1979). Failure of High Dose Vitamin C (ascorbic acid) therapy to benefit patients with advanced cancer. New. Eng. J. Med. 301: 687-690, 1979.

# Example: Vitamin C and Cancer

- Cameron and Pauling: not randomized; retrospective chart review
  - Raises doubts about comparability of groups
  - Doubts about equal treatment
- Creagan et al: randomized prospective study
  - Evidence that groups are comparable
  - Blinding reduces chance that groups are treated differently

# Cameron & Pauling (1976)

**ABSTRACT** Ascorbic acid metabolism is associated with a number of mechanisms known to be involved in host resistance to malignant disease. Cancer patients are significantly depleted of ascorbic acid, and in our opinion this demonstrable biochemical characteristic indicates a substantially increased requirement and utilization of this substance to potentiate these various host resistance factors.

The results of a clinical trial are presented in which 100 terminal cancer patients were given supplemental ascorbate as part of their routine management. Their progress is compared to that of 1000 similar patients treated identically, but who received no supplemental ascorbate.

The mean survival time is more than 4.2 times as great for the ascorbate subjects (more than 210 days) as for the controls (50 days) Analysis of the survival-time curves indicates that deaths occur for about 90% of the ascorbate-treated patients at one-third the rate for the controls and that the other 10% have a much greater survival time, averaging more than 20 times that for the controls.

The results clearly indicate that this simple and safe form of medication is of definite value in the treatment of patients with advanced cancer.

# Creagan et al. (1979)

ABSTRACT. 150 patients with advanced cancer participated in a controlled double blind study to evaluate the effects of high-dose vitamin C on symptoms and survival.

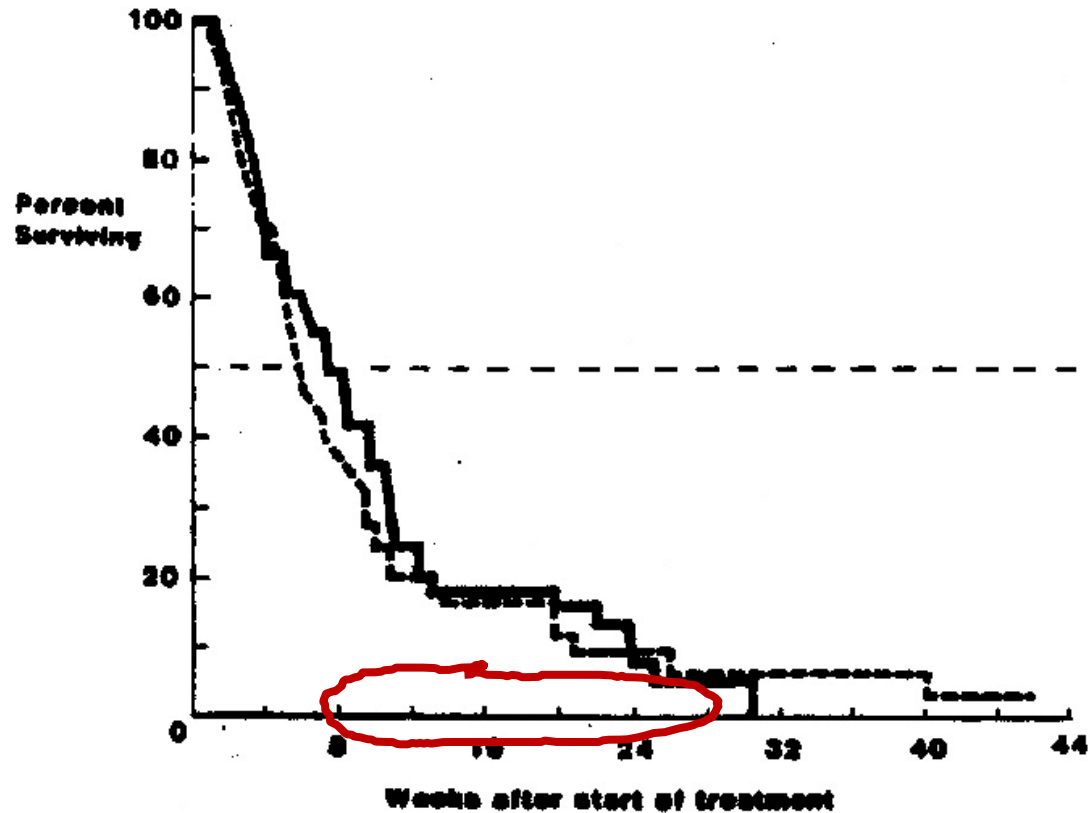
Patients were divided randomly into a group that received Vitamin C (10 g per day) and one that received a comparatively flavored lactose placebo. 60 evaluable patients received vitamin C and 63 received a placebo.

Both groups were similar in age, sex, type of primary tumor, performance score, tumor grade and previous chemotherapy.

The two groups showed no appreciable difference in changes of symptoms, performance status, appetite and weight. The median survival for all patients was about 7 weeks, and the survival times essentially overlapped.

In this selected group of patients, we were unable to show a therapeutic benefit of high-dose vitamin C.

# Kaplan-Meier Survival Curve



**Figure 1. High-Dose Vitamin C versus Placebo and Survival Results in Patients with Advanced Cancer.**

**The solid line shows survival in 60 patients given vitamin C. The dashed line shows survival in 63 patients given the lactose placebo.**

comparing treatments

# Conclusion

- Strengths and weaknesses of Cameron and Pauling?
- Strengths and weaknesses of Creagan et al.?
- Which result do you believe?
- Discuss in our zoom session