CLINICAL TRIALS WORKSHOP

SUNDAY NOVEMBER 12

NHMRC Clinical Trials Centre

Interpreting trial papers with more confidence

Kristy Robledo, Biostatistician

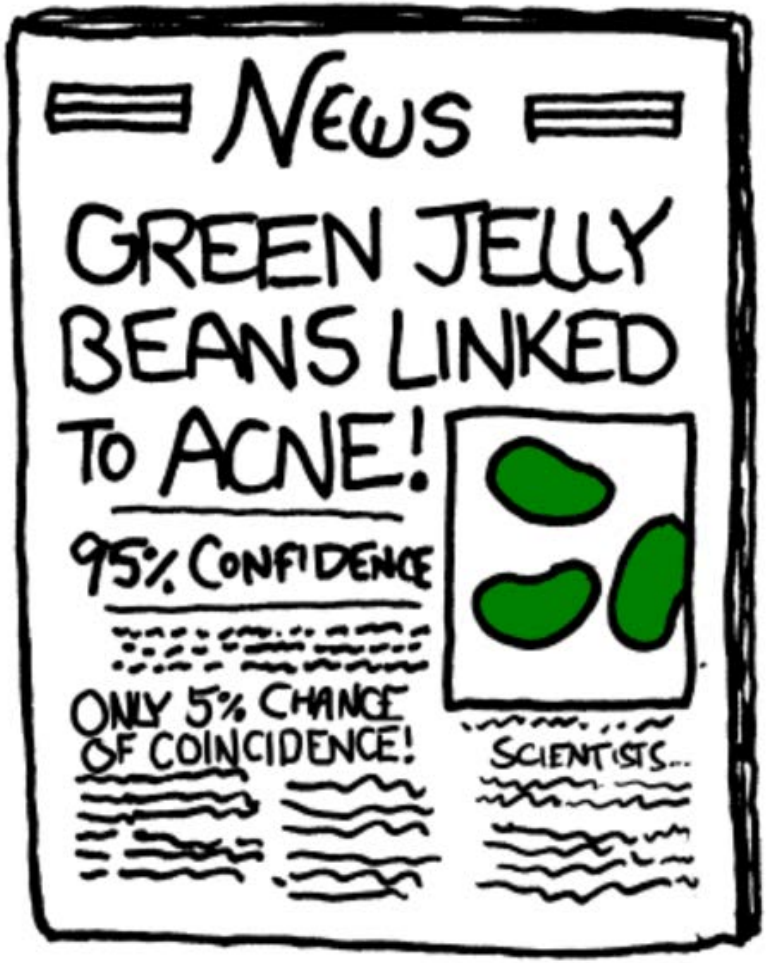
@redheadkristy

https://github.com/kristyrobledo/talks

THE UNIVERSITY OF SYDNEY

Aims

- Interpreting trials with more confidence!
  - Trial design
  - Primary endpoint
  - Sample size
  - Secondary outcomes
  - Kaplan-Meier curves
  - Subgroup analyses



NHMRC
Clinical Trials Centre

Trial Design

Design of clinical trials

Phase	I	II	III
Does it?	Hurt?	Help?	Help much?
Is it …. ?	Safe?	Active?	Effective?
No of patients	~10-30	~30-100	~100-300
Pluses	Early access	Promising	Definitely active
Minuses	Safety unknown, toxic??	Phase I could have been wrong?	Only 50:50 chance to get new drug
Drug of name:	XYZ-123	Unpronounceable-a-mab	Drugzilla™
Locations	One centre	Few hospitals	World-wide

Controlled trials

			Comparison of Exp v Con
Strong Control	Parallel	Exp Con	Between subjects
	Crossover	Exp → Con Con → Exp	Within subjects
Weak/No Control	Before-after	Con → Exp	Within subjects
	Single arm	Exp	?

Design… depends on the Q

- Safety of drug?
  - Not yet used in humans?
  - Used in other cancers?
- Comparing groups?
  - For treatment better than control?
  - For treatment no worse than control?

Phase 1

Phase 2

Phase 3

Superiority

Non-inferiority

Example: Phase II

VOLUME 34 · NUMBER 19 · JULY 1, 2016

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Response to Cetuximab With or Without Irinotecan in Patients With Refractory Metastatic Colorectal Cancer Harboring the *KRAS* G13D Mutation: Australasian Gastro-Intestinal Trials Group ICECREAM Study

Eva Segelov, Subotheni Thavaneswaran, Paul M. Waring, Jayesh Desai, Kristy P. Robledo, Val J. Gebski, Elena Elez, Louise M. Nott, Christos S. Karapetis, Sebastian Lunke, Lorraine A. Chantrill, Nick Pavlakis, Mustafa Khasraw, Craig Underhill, Fortunato Ciardiello, Michael Jefford, Harpreet Wasan, Andrew Haydon, Timothy J. Price, Guy van Hazel, Kate Wilson, John Simes, and Jeremy D. Shapiro

Example: Phase III

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

A Phase 3 Randomized Trial of Nicotinamide for Skin-Cancer Chemoprevention

Andrew C. Chen, M.B., B.S., Andrew J. Martin, Ph.D., Bonita Choy, M.Med., Pablo Fernández-Peñas, Ph.D., Robyn A. Dalziell, Ph.D., Catriona A. McKenzie, M.B., B.S., Richard A. Scolyer, M.D., Haryana M. Dhillon, Ph.D., Janette L. Vardy, M.D., Anne Kricker, Ph.D., Gayathri St. George, M.Sc.Med., Niranthari Chinniah, M.B., B.S., Gary M. Halliday, D.Sc., and Diona L. Damian, Ph.D.

In the next 5 minutes….

- How would you describe the trial design for:
  - For ICECREAM?
  - For ONTRAC?

In the next 5 minutes….

- How would you describe the trial design:
  - For ICECREAM?
    Phase II, parallel arm, randomised controlled trial
  - For ONTRAC?
    Phase III, double blind, parallel arm, randomised controlled trial

NHMRC
Clinical Trials Centre

Primary endpoint

Endpoints

- Many different types of endpoints exist…
  - Continuous eg. waist measurements, units of blood used in emergency procedures
  - Binary eg. cured/not cured, alive/dead
  - Ordinal/nominal eg. strongly disagree/disagree/neutral/ agree/strongly agree ;

    or prefer tablet/prefer lotion/prefer injection
  - Count (eg. number of asthma attacks per week, number of standard drinks consumed in the last 24 hours)
  - Time-to-event (eg. time taken to eradicate symptoms, time taken to achieve pregnancy)

Outcomes and endpoints

- Outcomes should correspond with objectives and hypotheses
- Outcome measures should be reliable and valid
  - Clinical assessment (with adjudication)
  - Self-administered questionnaires
- **The primary endpoint**
  - Should be pre-specified
  - Most compelling – what would convince the sceptics
  - Choose to believe above all others – results often conflict
  - Robust, transparent, valid, measureable on all patients!
  - Influences all aspects of design (***including sample size***)
  - Blinded assessment preferable

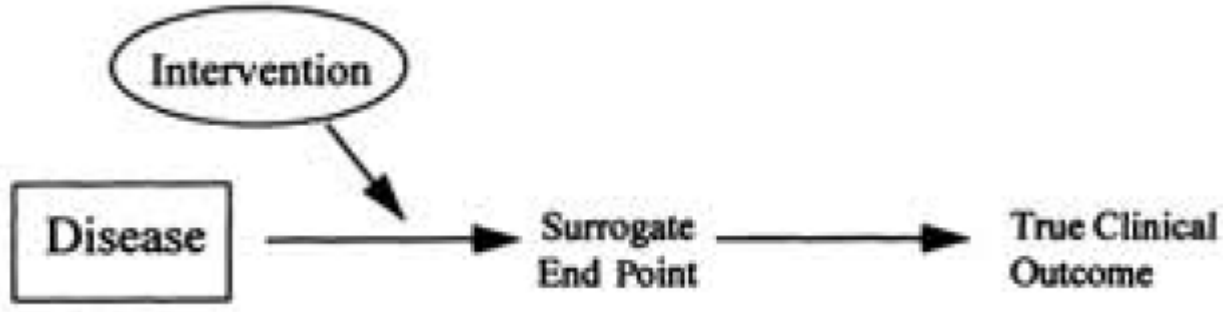
Endpoints

- It is usual practice to base the sample size calculation around the PRIMARY endpoint
  - the main outcome the study is using to answer an important clinical question.

- Not uncommon to have other secondary endpoints collected
  - Allow investigation of subsidiary questions that do not have the same priority as the primary objective of the trial

Definitive vs surrogate endpoints

- Definitive endpoint
  - Relevant to the patient and clinically meaningful (e.g. mortality, ?quality of life, etc.)
- Surrogate endpoint
  - A substitute for the definitive endpoint
    - E.g. tumour response as a surrogate for survival time
  - The effect of the intervention on the surrogate should predict the effect on the definitive endpoint. Evidence of a correlation alone is insufficient.



Ann Intern Med. 1996;125:605-613.

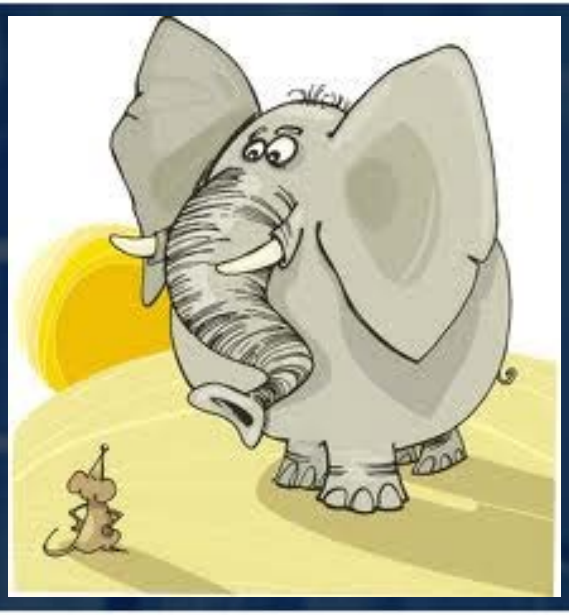
In the next 5 minutes….

- What are the primary endpoints and the results:
  - For ICECREAM?
  - For ONTRAC?

In the next 5 minutes….

- What are the primary endpoints and the results:
  - For ICECREAM?
    % progression free at 6 mths
    C = 10% (2-26%), C+I = 23% (9-40%)
    HR = 0.74 (0.41-1.13)
  - For ONTRAC?
    New nonmelanoma skin cancers in 12 mths
    C = 2.4pa, N = 1.8pa
    Diff in rates = 27% (5-44%)

Sample size



How large does an RCT need to be?

- Depends on:
  1. Anticipated size of the treatment effect
  2. How sure we want to be about the conclusions we draw from the data collected

     (constraining errors)

Non compliance

- Patients not adhering to their randomised intervention, will dilute the effect between the control or standard intervention and the new intervention
- Moving from active to control (drop-out of active)
  - E.g. in a trial evaluating an exercise program on mental health, a participant from the active exercise group might stop exercising
- Moving from control to active (drop-in to active)
  - E.g. in a trial evaluating an exercise program on mental health, someone begins exercising despite being randomised to the control group

Consequences of inadequate SS

- Scientific merit
  - Inconclusive results
  - Unreliable or misleading results
- Risks
  - An effective treatment may appear inferior due to the play of chance (small studies can be risky to the adoption of effective therapies)
  - Waste of resources
- Ethics
  - Patient participation in studies of limited value
  - IRB may not approve such studies

In the next 5 minutes….

- What are the sample size calculations:
  - For ICECREAM?
  - For ONTRAC?

ICECREAM

Statistical assumptions were on the basis of retrospective data with a median PFS of 1.8 months in the cetuximab-alone arm versus 4.0 months for patients with *KRAS* G13D–mutated tumors receiving any cetuximab therapy.[5] This corresponds to a 6-month PFS rate of approximately 10% for the monotherapy group and 35% for the combination group. The G13D component of the study planned to recruit 25 patients per arm to match the wild-type cohort. This enabled 80% power ($\alpha = .05$) to detect an improvement from 15% to 40% in 6-month PFS by using the Simon design for phase II trials. This magnitude of effect was chosen because we were interested in detecting a

ONTRAC

NHMRC Clinical Trials Centre

STATISTICAL ANALYSIS

We estimated that with a sample size of 386, the study would have 90% power to detect a 33% lower rate of new nonmelanoma skin cancers with nicotinamide than with placebo at 12 months at a 5% level of significance, assuming that non-melanoma skin cancer counts would follow a Poisson distribution and that a mean of 1.0 new nonmelanoma skin cancers per person would be detected in the placebo group, and allowing for an average rate of nonadherence of up to 10%.

NHMRC
Clinical Trials Centre

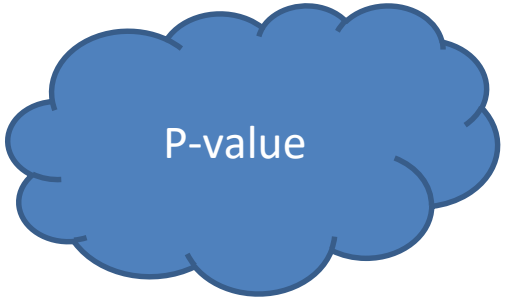
Secondary outcomes

Analysis of endpoints

- **Analysis of binary endpoints - Proportions**
  - Difference in proportions
  - Odds ratios
  - Relative risks

- **Analysis of continuous endpoints**
  - Difference in means

- **Analysis of Time to event (TTE) endpoints**
  - Graphical displays (KM plots)
  - Log-rank test
  - Proportional hazards regression



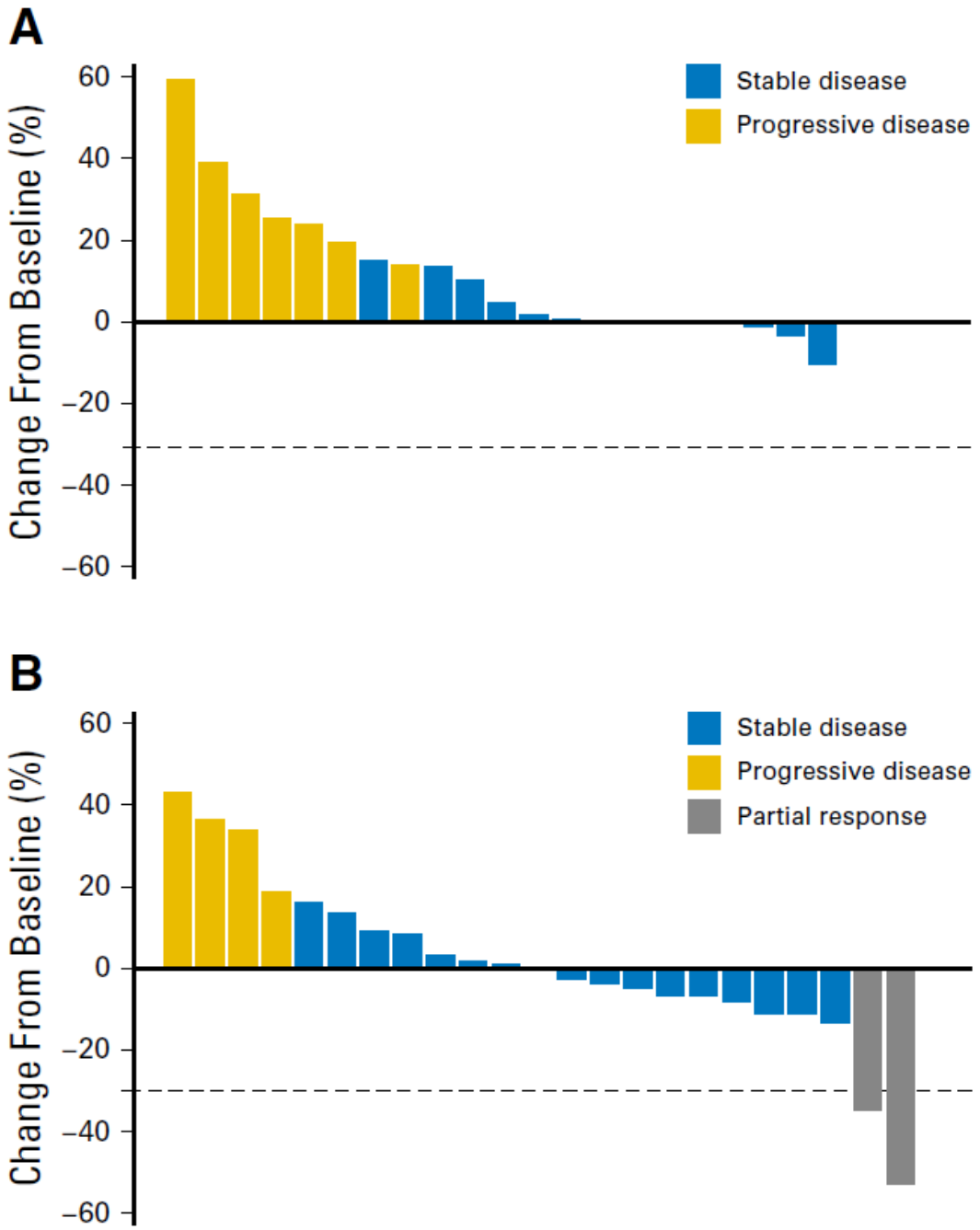
What are the ICECREAM secondary outcomes?

- Response rate (complete or partial)
- Overall survival
- QOL
  - FACT – C
  - DLQi
  - FACT-EGRFI-18

ICECREAM: response

Secondary End Points

No responses were seen in the cetuximab arm and 58% of patients had stable disease. In the cetuximab-irinotecan arm, 9% of patients had a partial response and 70% had stable disease. The best response by treatment arm is shown in Figures 3A and 3B. A post



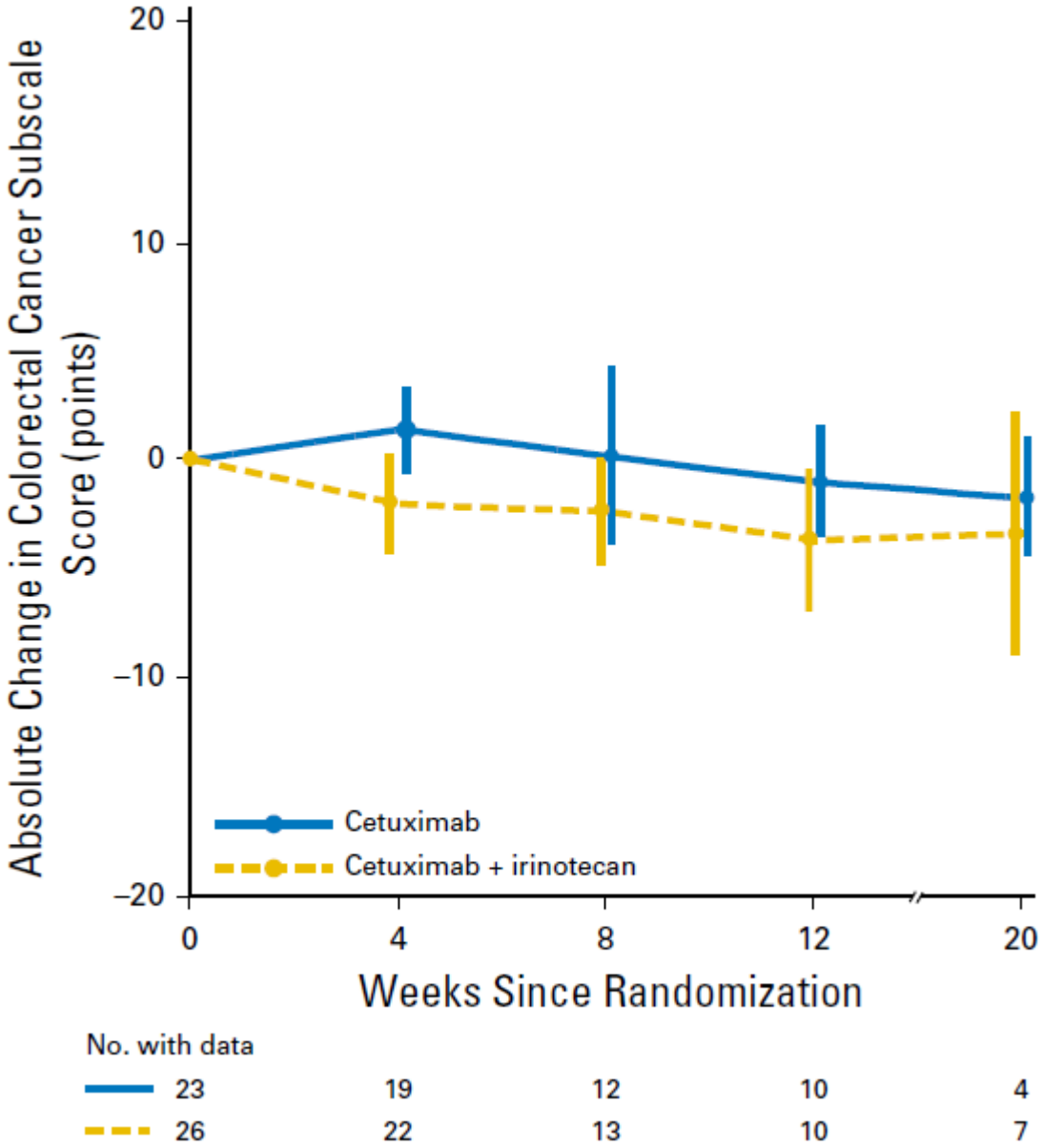
Cetux

Cetux +Irino

QOL

because of clinician or patient preference. Regarding QoL, the average CRC subscale at 20 weeks was 0.43 points (95% CI, –0.9 to 1.8 points) poorer compared with baseline for the cetuximab arm and 2.6 points (95% CI, –0.2 to 5.5 points) poorer for the cetuximab-irinotecan arm, with no significant difference between study arms (Fig 5). The average skin QoL (measured by the Dermatology Life Quality Index) at 12 weeks was 3.6 points (95% CI, –0.8 to 8.0 points) poorer for the cetuximab arm and 1.4 points (95% CI, –0.4 to 2.4 points) poorer for the cetuximab-irinotecan arm, with no difference between treatments. Similar results were seen for FACT-EGFRI-18.

QOL

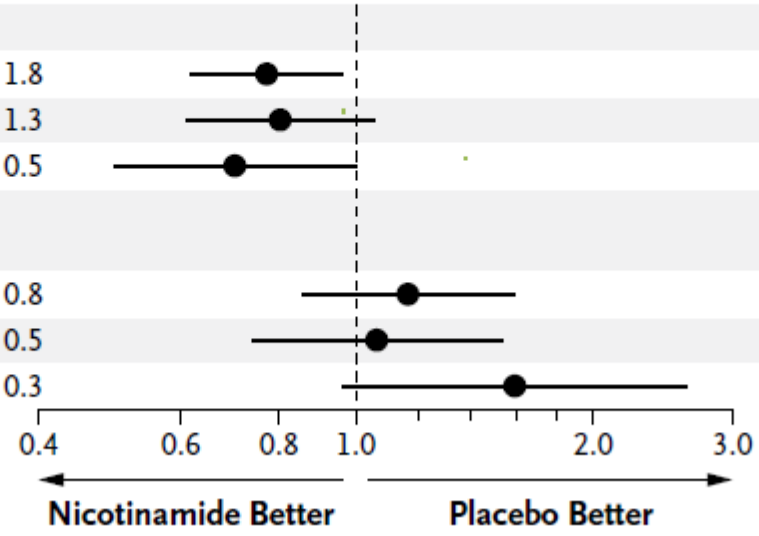

Absolute Change in Colorectal Cancer Subscale Score (points)
20
10
0
−10
−20
Cetuximab
Cetuximab + irinotecan
0
4
8
12
20
Weeks Since Randomization
No. with data
23 19 12 10 4
26 22 13 10 7

What are the ONTRAC secondary outcomes?

- New basal cell carcinomas, squamous cell carcinomas and actinic keroses in 12 mths
- New NMSC in 6 mths
- Safety of nicotinamide
- Cognitive function
- Transepidermal water loss

ONTRAC

Subgroup	Placebo (mean no. of lesions/person)	Nicotinamide (mean no. of lesions/person)	Relative Difference, % (95% CI)	P Value
12-mo intervention period				
NMSCs	2.4	1.8	23 (4 to 38)	0.02
BCCs	1.7	1.3	20 (–6 to 39)	0.12
SCCs	0.7	0.5	30 (0 to 51)	0.05
6-mo postintervention period				
NMSCs	0.8	0.8	–17 (–59 to 14)	0.33
BCCs	0.6	0.5	–6 (–53 to 26)	0.73
SCCs	0.3	0.3	–59 (–163 to 4)	0.07



ONTRAC: SAFETY

No clinically significant be
ences were found with respe
types of adverse events th
study groups (Table S3 in the
pendix). The terms for the

curred in the two groups combined included neoplasm (12 patients), cardiac chest pain (9 patients), fall (7 patients), lung infection (6 patients), atrial fibrillation (6 patients), injury (6 patients), heart failure (5 patients), and hematoma (5 patients). Two internal cancers were diagnosed in the placebo group (duodenal carcinoma diagnosed at month 1 of the study and lung cancer at month 3) and five internal cancers were diagnosed in the nicotinamide group (non-Hodgkin's lymphoma diagnosed at month 1 of the study, colorectal cancer at month 2, lung cancer at month 2, prostate cancer at month 7, and bladder cancer at month 9). Four new invasive melanomas and six new melanomas in situ were diagnosed during the 12-month intervention period and were evenly distributed between the two groups. A microcystic adnexal carcinoma

NHMRC
Clinical Trials Centre

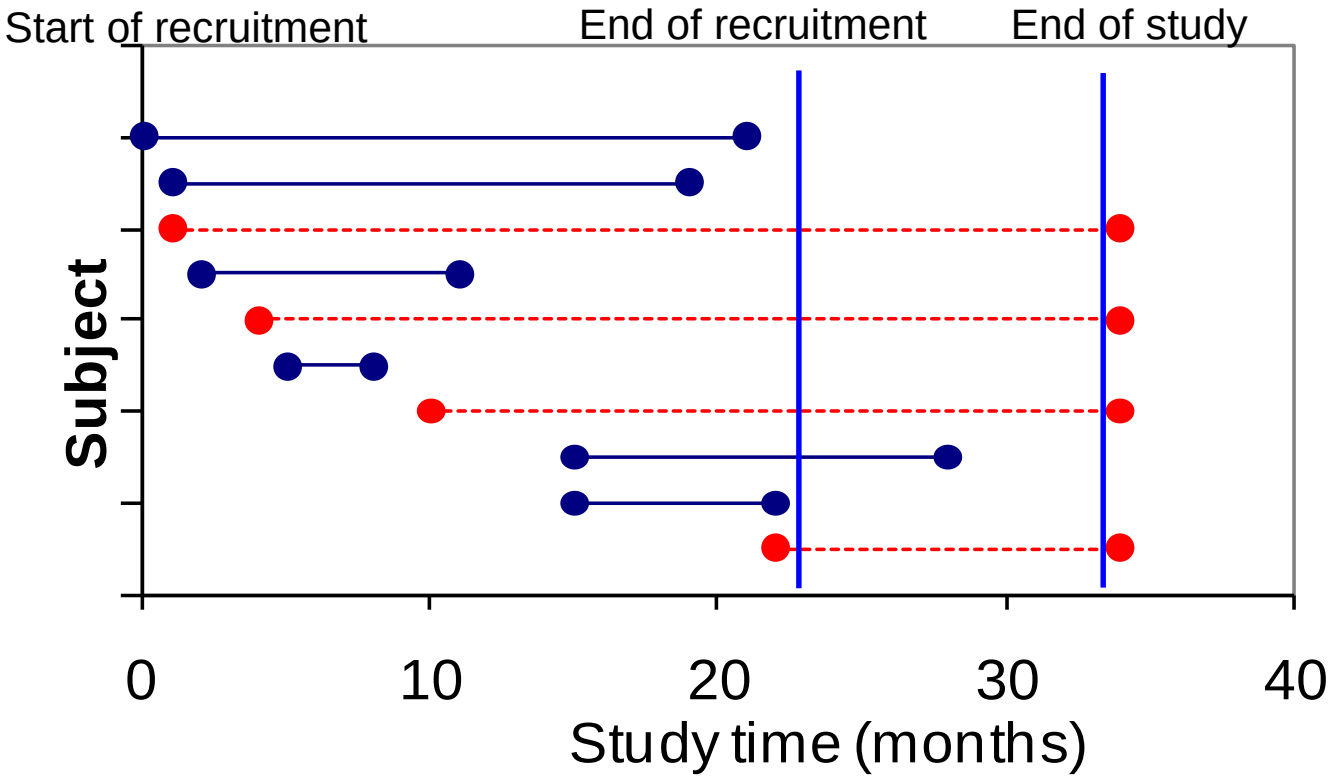
Kaplan-Meier curves or time to event data

What is time to event data?

- Any data measuring time from a well defined time origin to a well defined endpoint is referred to as '**time to event**' or '**survival**' data.
  - *Time origins*: diagnosis of a particular condition, start of a treatment regime.
  - *Endpoints*: death, relief from pain, recurrence of symptoms, birth, tumour progression.
- Note that the endpoint need not necessarily be 'death' and that it need not necessarily be a negative experience, as the term *survival* may imply.
- Time to event data are usually continuous and often right skewed.

Time to event data: Example

- In a study:
  - 10 patients enter, the first on Day 0 and the 10th at month 22.
  - 6 experience the event of interest before the end of the study, 4 do not.



What is censoring?

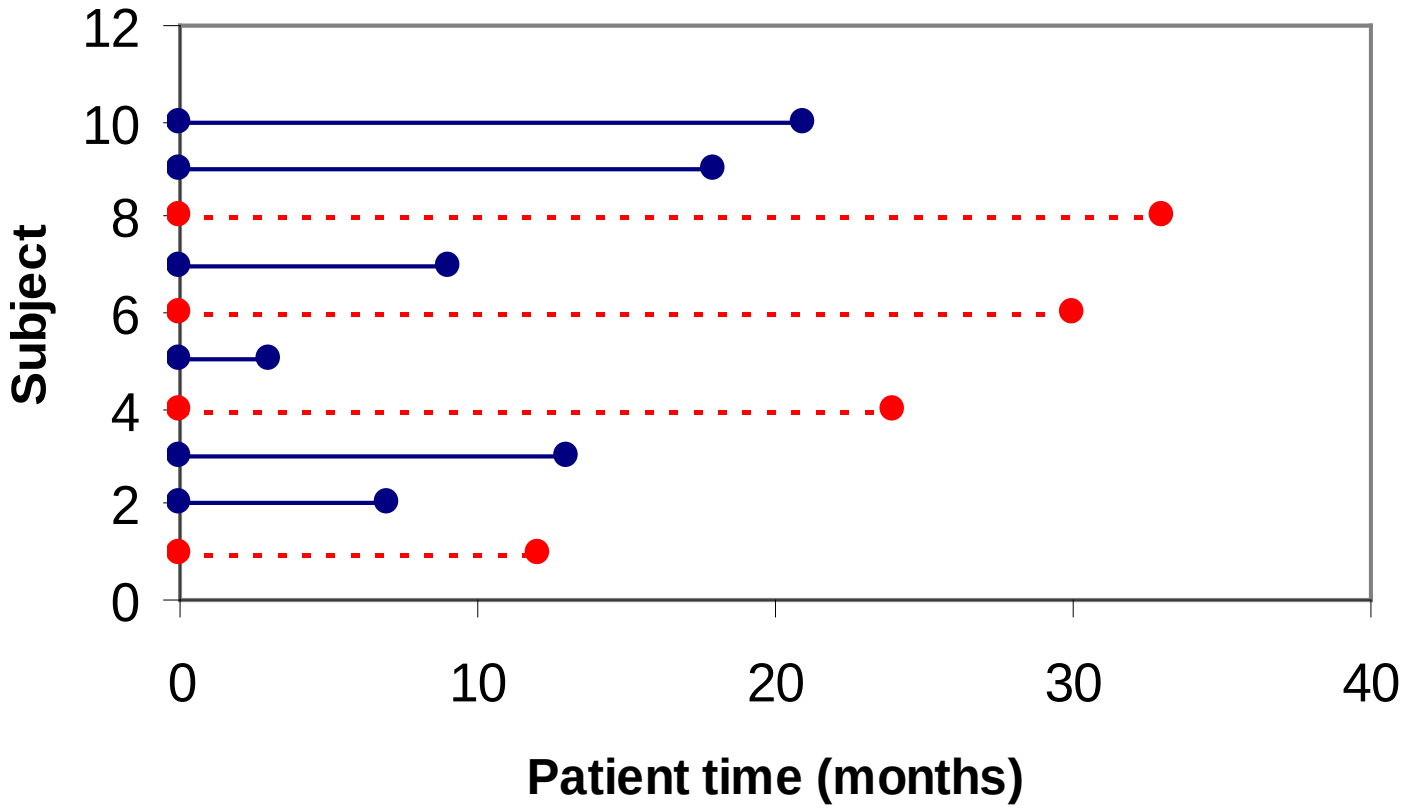
- For those that do not have the event of interest, we know only that their 'survival time' was longer than the time for which they have been observed.
- Example:
  - In a clinical trial, monitoring time from relapse to death, not all patients will necessarily die during the follow-up period.

These incomplete observations are known as

'**censored observations**'

Time to event data: Example

- From the study subject's point of view, the time will look like:



Dotted line = censored observations
Solid line = uncensored observations

Time to event analysis

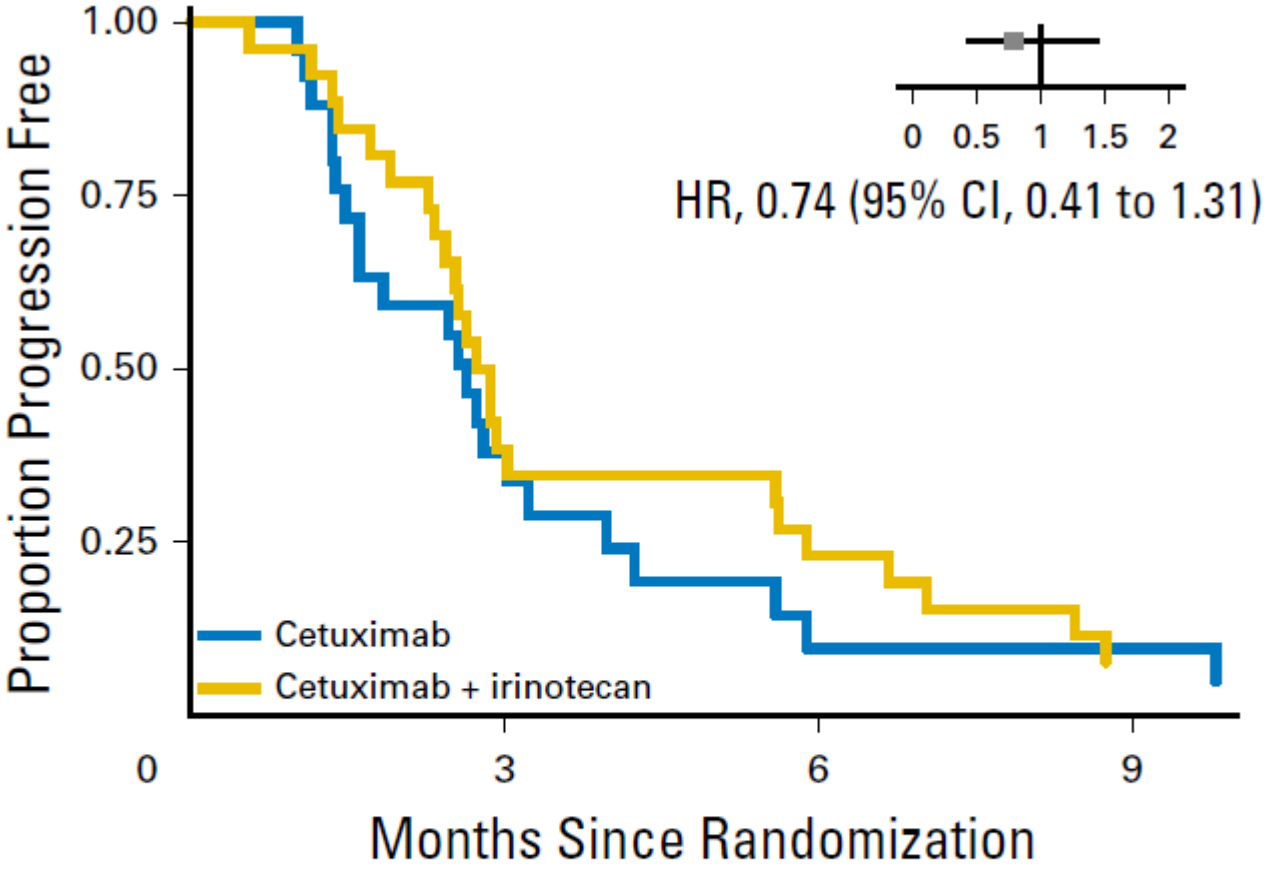
- Time to event analysis handles both censored and uncensored data, making use of *all* the available information.

- It is used to:
  - Describe and graphically present time to event data.
  - Test for differences in the time to event of different groups of subjects.
  - Perform regression modelling of time to event data.

In the next 5 minutes….

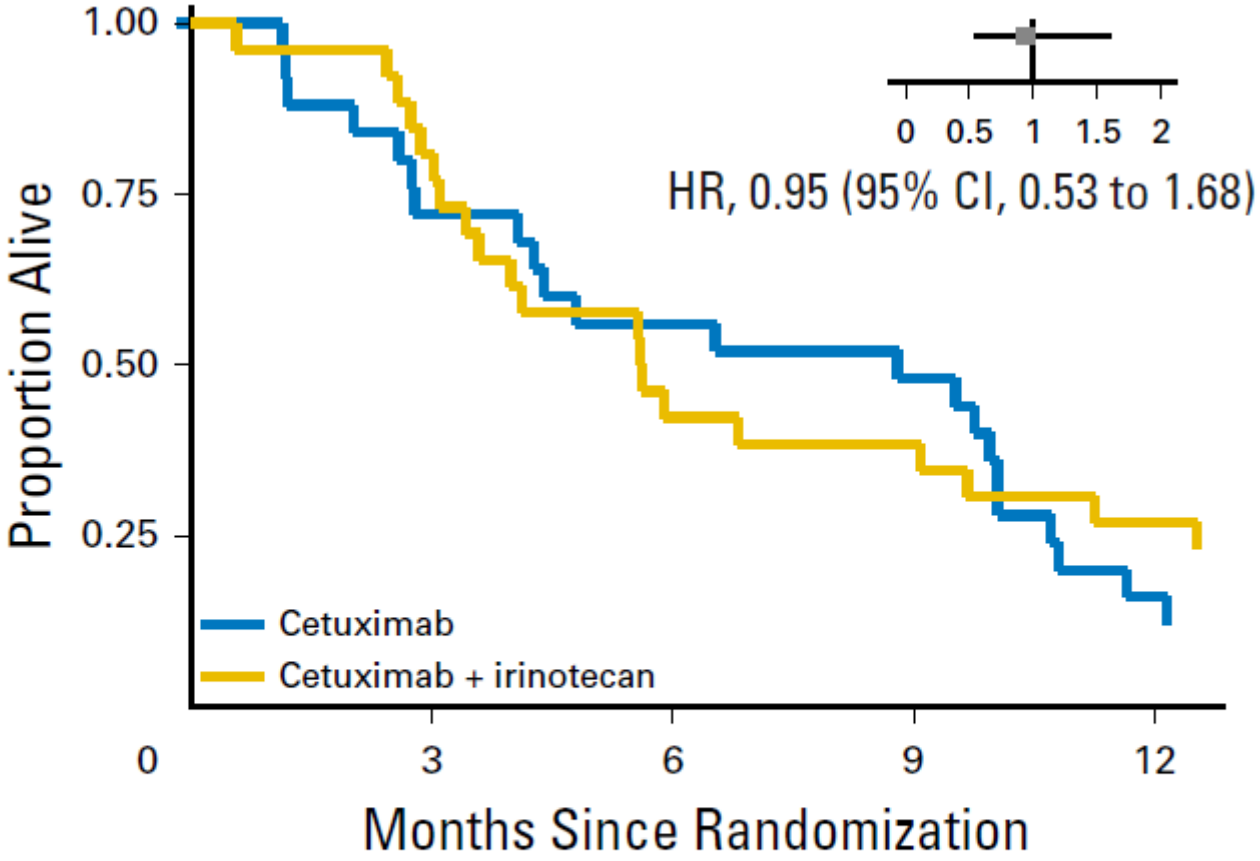
- Identify the Kaplan-Meier curves and any Hazard ratios reported:
  - For ICECREAM?
  - For ONTRAC?

ICECREAM: PFS



No. at risk				
Cetuximab	25	9	2	2
Cetuximab + irinotecan	26	10	6	2

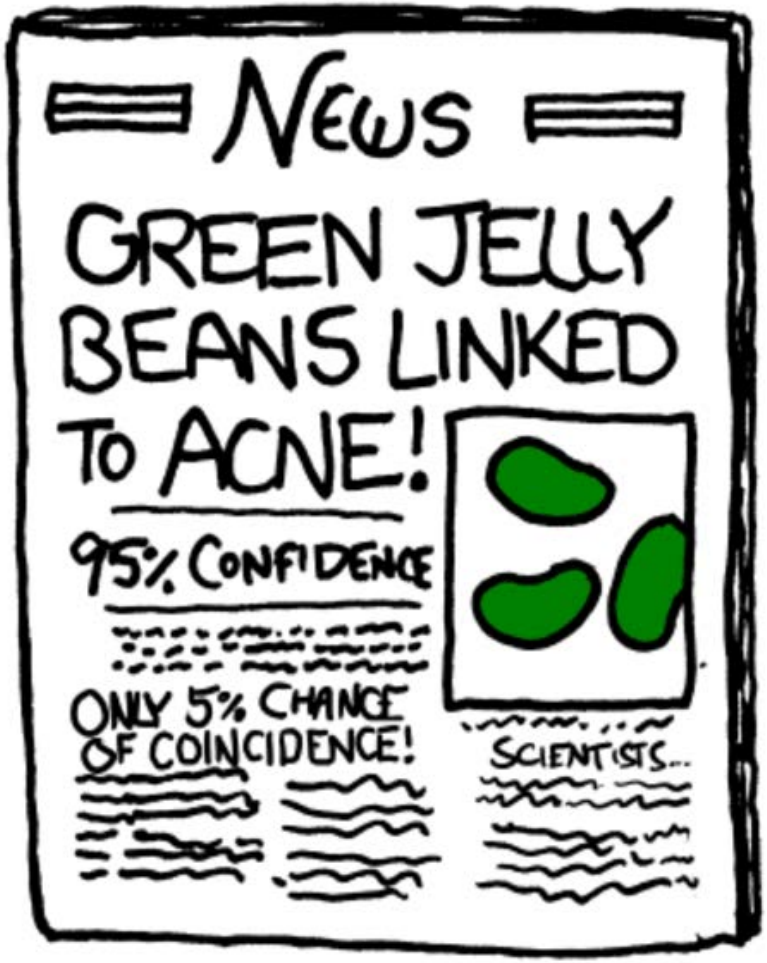
ICECREAM: OS



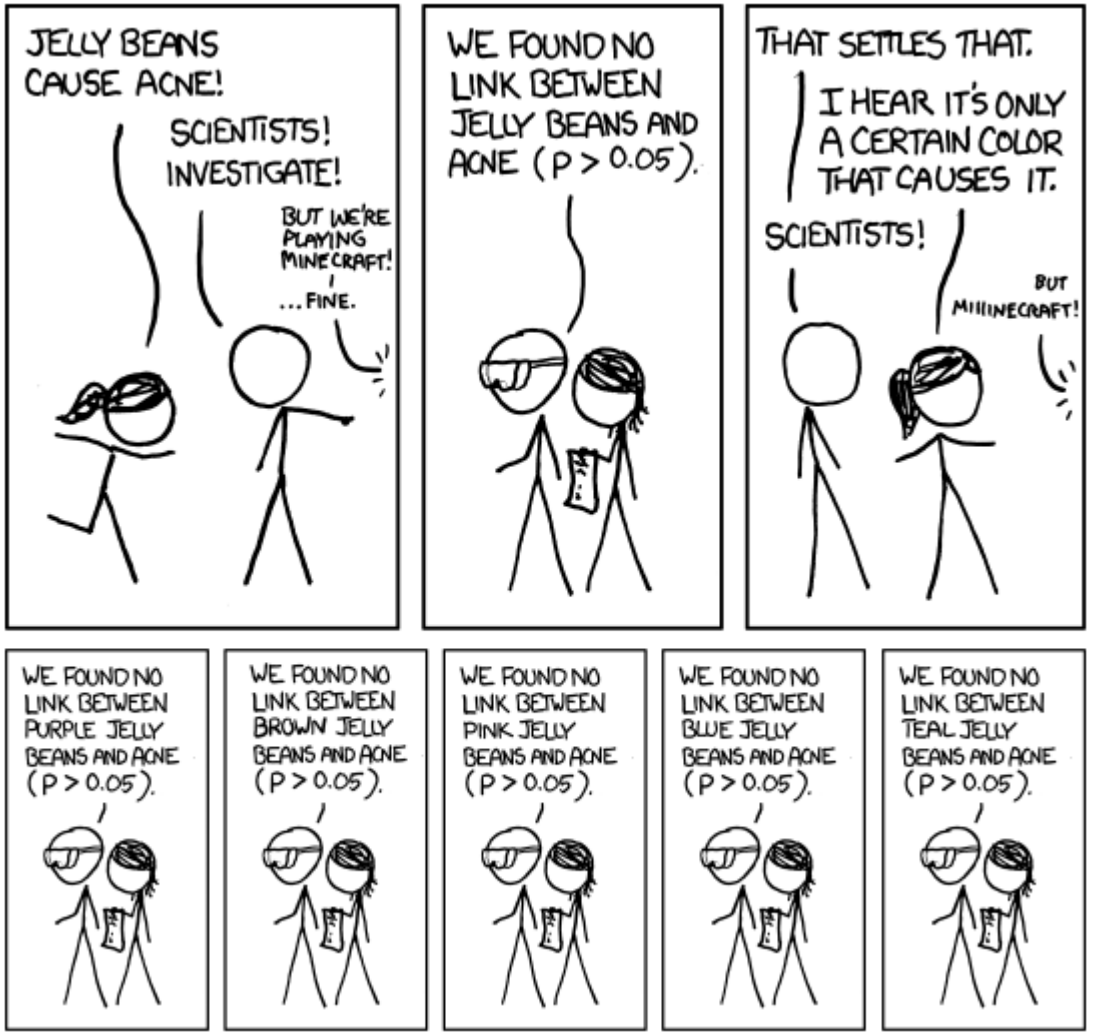
No. at risk					
Cetuximab	25	21	14	13	5
Cetuximab + irinotecan	26	21	11	10	7

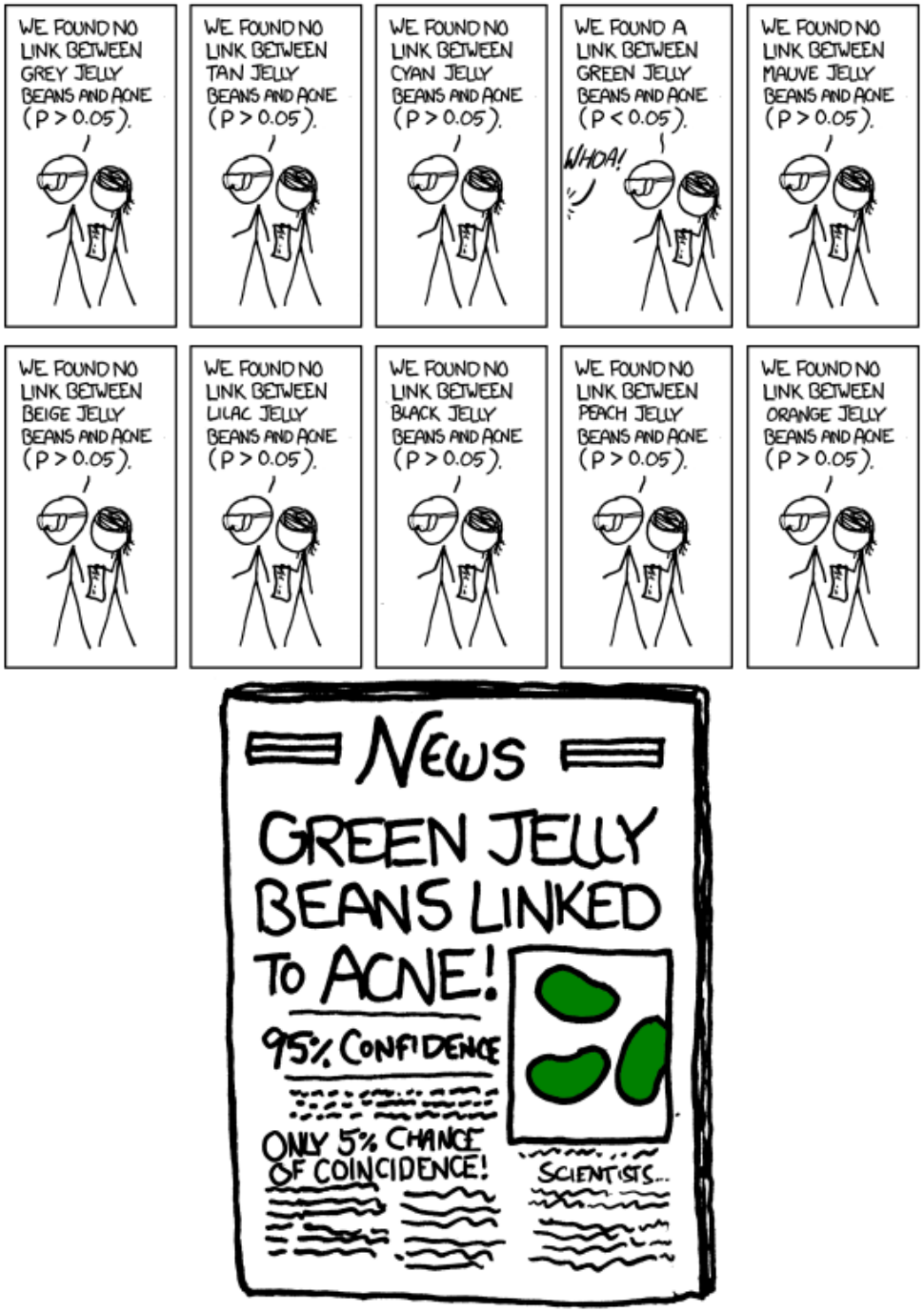
Subgroup analyses

How do we interpret this result?



Now in context….




WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

The multiple comparisons problem

- Under the null hypothesis of *no effect*, we would expect 5% of (independent) hypothesis tests to produce a statistically significant result (i.e. a p-value <5%)

Number of Independent Tests (k) Performed at α=0.05	**Probability at Least One Significant Result $1 – (1-\alpha)^k$**
1	0.05
2	0.10
3	0.14
5	0.23
10	0.40
25	0.72

Example

Diabetologia (2001) 44: 312–319

Diabetologia

(Very) weak evidence that …

Substituting dietary saturated for monounsaturated fat impairs insulin sensitivity in healthy men and women: The KANWU study

B. Vessby[1], M. Uusitupa[2], K. Hermansen[3], G. Riccardi[4], A. A. Rivellese[4], L. C. Tapsell[5], C. Nälsén[1], L. Berglund[1], A. Louheranta[2], B. M. Rasmussen[3], G. D. Calvert[5], A. Maffetone[4], E. Pedersen[3], I.-B. Gustafsson[1], L. H. Storlien[5]

[1] Unit for Clinical Nutrition Research, Department of Public Health and Caring Sciences/Geriatrics, University of Uppsala, Uppsala, Sweden
[2] Department of Clinical Nutrition, University of Kuopio, Kuopio, Finland
[3] Department of Clinical Endocrinology and Metabolism, Aarhus Amtssygehus and Aarhus University, Aarhus, Denmark
[4] Department of Clinical and Experimental Medicine, School of Medicine, Federico II University, Naples, Italy
[5] Department of Biomedical Sciences and Medical Research Unit, University of Wollongong, Wollongong, Australia

Design: N=162 healthy subjects randomly assigned to receive
(i) high SFA or
(ii) high MUFA diet for 3 months

Example

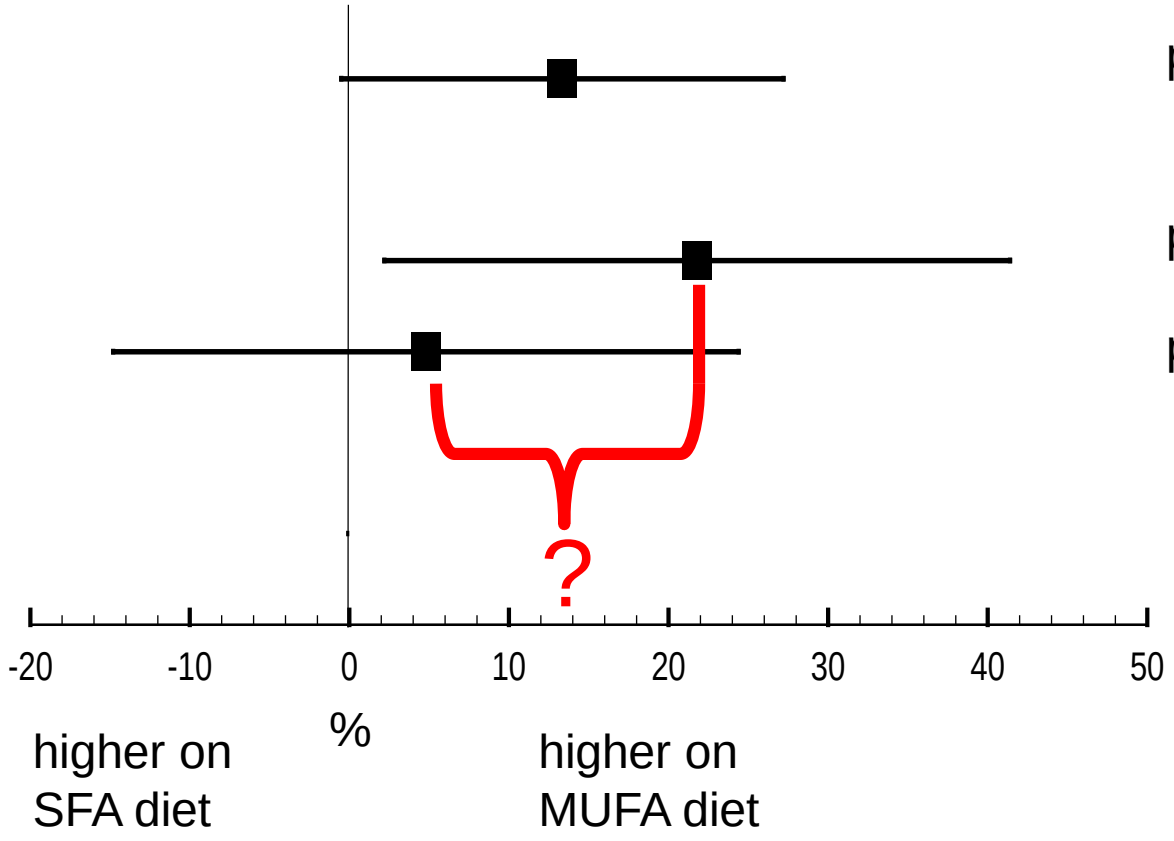
Effect of diet (MUFA v SFA) on insulin sensitivity with 95% CI

MAIN RESULT

Subgroup: Fat intake

Low

High



How to analyse subgroups

- **Formal test of interaction**
  - Direct comparison between estimated effect in each subgroup

Effect for low fat group = 22%

Effect for high fat group = 5%

Difference between subgroups = 17% (95% CI: -11% to 45%)

NHMRC Clinical Trials Centre

Effect of diet (MUFA v SFA) on insulin sensitivity with 95% CI

MAIN RESULT

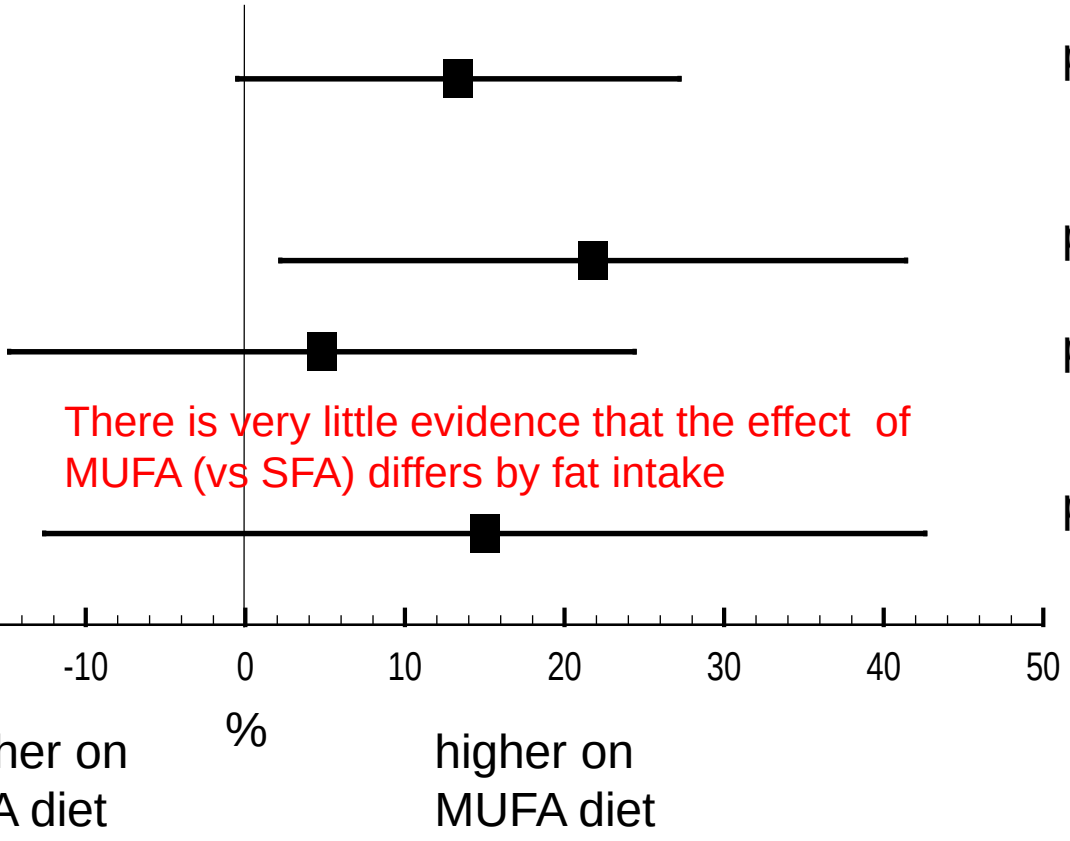
Subgroup: Fat intake

Low

High

Formal test of interaction

Effect(Low) – Effect(High)



Principles for subgroup analysis

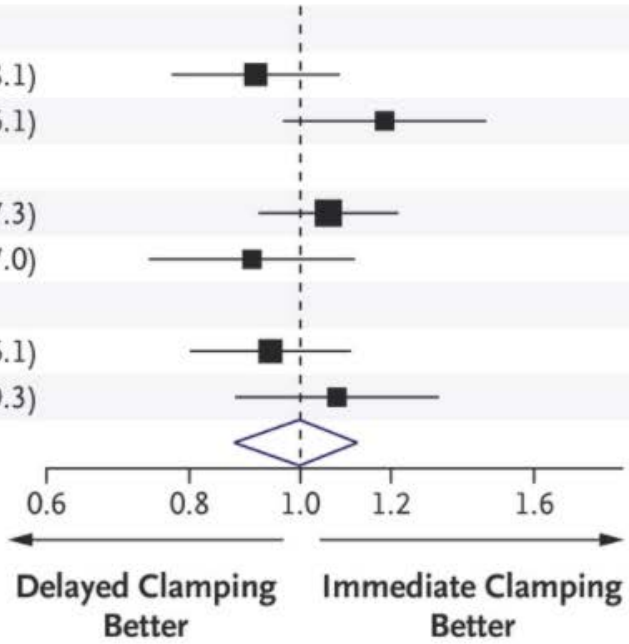
- Subgroup analyses should be:
  - Pre-specified in the trial protocol and limited to a small number of biologically plausible questions that are supported by evidence from other independent studies
  - Conducted using appropriate statistical tests
  - Fully reported (i.e. all subgroup analyses performed should be reported)
  - Evaluated in the context of a multiple comparisons problem
- When uncertainty remains about reliability of results, investigators should attempt to confirm using data from other RCTs (ideally a pooled analysis)

In the next 5 minutes….

- Are there any subgroup analyses:
  - For ICECREAM?
  - For ONTRAC?

In another NEJM publication….

Subgroup	Delayed Cord Clamping	Immediate Cord Clamping	Relative Risk of Death or Major Morbidity by 36 Wk of Postmenstrual Age (95% CI)	P Value for Interaction
	no. of events/total no. (%)			
Sex				0.10
Male	146/413 (35.4)	166/436 (38.1)	0.91 (0.77–1.08)	
Female	131/335 (39.1)	113/313 (36.1)	1.18 (0.97–1.45)	
Gestational age				0.23
<27 wk	160/265 (60.4)	145/253 (57.3)	1.06 (0.92–1.22)	
≥27 wk	117/483 (24.2)	134/496 (27.0)	0.91 (0.74–1.12)	
Type of delivery				0.23
Cesarean	166/492 (33.7)	176/487 (36.1)	0.94 (0.80–1.11)	
Vaginal	111/256 (43.4)	103/262 (39.3)	1.08 (0.88–1.32)	
Overall			1.00 (0.88–1.13)	



Take home messages

- Trial design
  - Fed from the study question
- Primary endpoint
  - Main study question and the sample size
  - Relevant, convincing
  - Determines positive/negative study
- Sample size
  - How sure do we want to be
  - Anticipated size of the treatment effect

Take home messages

- Secondary outcomes & Subgroups
  - Hypothesis generating
  - Interpret with caution!
- Kaplan-Meier curves
  - Report extent of FU (# at risk)
  - Be careful with data to the right
  - Effect size and 95%CI
  - Don’t over interpret shapes of curves (beware of squarish lines)