The challenges of a systematic review and meta-analysis of prognosis studies

Richard D. Riley [1]

Willi Sauerbrei [2] , Douglas G. Altman [3]

1 Centre for Medical Statistics & Health Evaluation, University of Liverpool

2 Institute of Medical Biometry and Medical Informatics, University Hospital of Freiburg

3 Centre for Statistics in Medicine, University of Oxford

Funding sources

UK HTA Programme

DH NCCRCD

Aims of this Talk

- Examine if a systematic review and meta-analysis is feasible for prognostic marker studies

- **Highlight major problems; including**:

  Poor reporting of results in primary studies

  Heterogeneity across studies

  Selective reporting / publication bias

- Consider guidelines and approaches to limit these problems

- Encourage the availability of individual patient data

- Consider reasons to be optimistic

Prognostic Markers

- Also called ***prognostic variables or factors***

- Identify different risk groups
  - help to stratify patients for treatment
  - ensure balanced groups within RCTs
  - aid patient counselling

- Include biological, clinical, genetic, histological, pathological and demographic features.

- Example: CEA in colorectal cancer
  MYCN in neuroblastoma
  Age in traumatic brain injury

Evidence-Based Prognostic Markers

- Primary studies of prognostic markers important

- Clinical use of markers ideally based on overall evidence

  This is difficult for clinicians because:

  - Large number of primary studies
  - Conflicting results
  - Small patient numbers

- Formal evidence-based reviews and synthesis of prognostic marker studies needed

Systematic Reviews and Meta-analysis

- **Systematic Reviews**
  - common approach (e.g. Cochrane)
  - identifying, evaluating & combining evidence-base
  - systematic & transparent framework

- **Meta-analysis**

- statistical analysis
- combines quantitative results across studies
- produces overall summary of effect of interest
- increase power, reduce uncertainty
- can examine impact of study-level covariates

Meta-analysis using aggregate data

- Traditional meta-analysis uses aggregate data

- Obtainable from publications or study authors

- Meta-analysis of prognostic marker studies usually requires *from each study*:

  - **an estimate** of the relationship between the marker and outcome;

    e.g. hazard ratio for overall survival

  - **the standard error** of this estimate;

    e.g. standard error of log hazard ratio

- Meta-analysis synthesises the results

  e.g. each study weighted by inverse of the variance

Example of a meta-analysis

Is VMA a prognostic marker for overall survival in neuroblastoma?

Is a Systematic Review and Meta-Analysis of Prognostic Markers in Neuroblastoma Possible?

- **Neuroblastoma**
  - most common solid tumour of childhood
  - active research area for prognostic markers

- **Prognostic Tumour Markers**
  - Measurable parameter in the blood, urine or body tissue e.g. CEA (protein), Chromosome 1p (gene).

- **Systematic Review** of primary studies reporting results for a potential prognostic tumour marker neuroblastoma

- ‘**A Prognosis Paper**’: one presenting aggregate data or individual patient data (IPD) relating marker levels at baseline to survival

Identifying the Prognostic Marker Literature

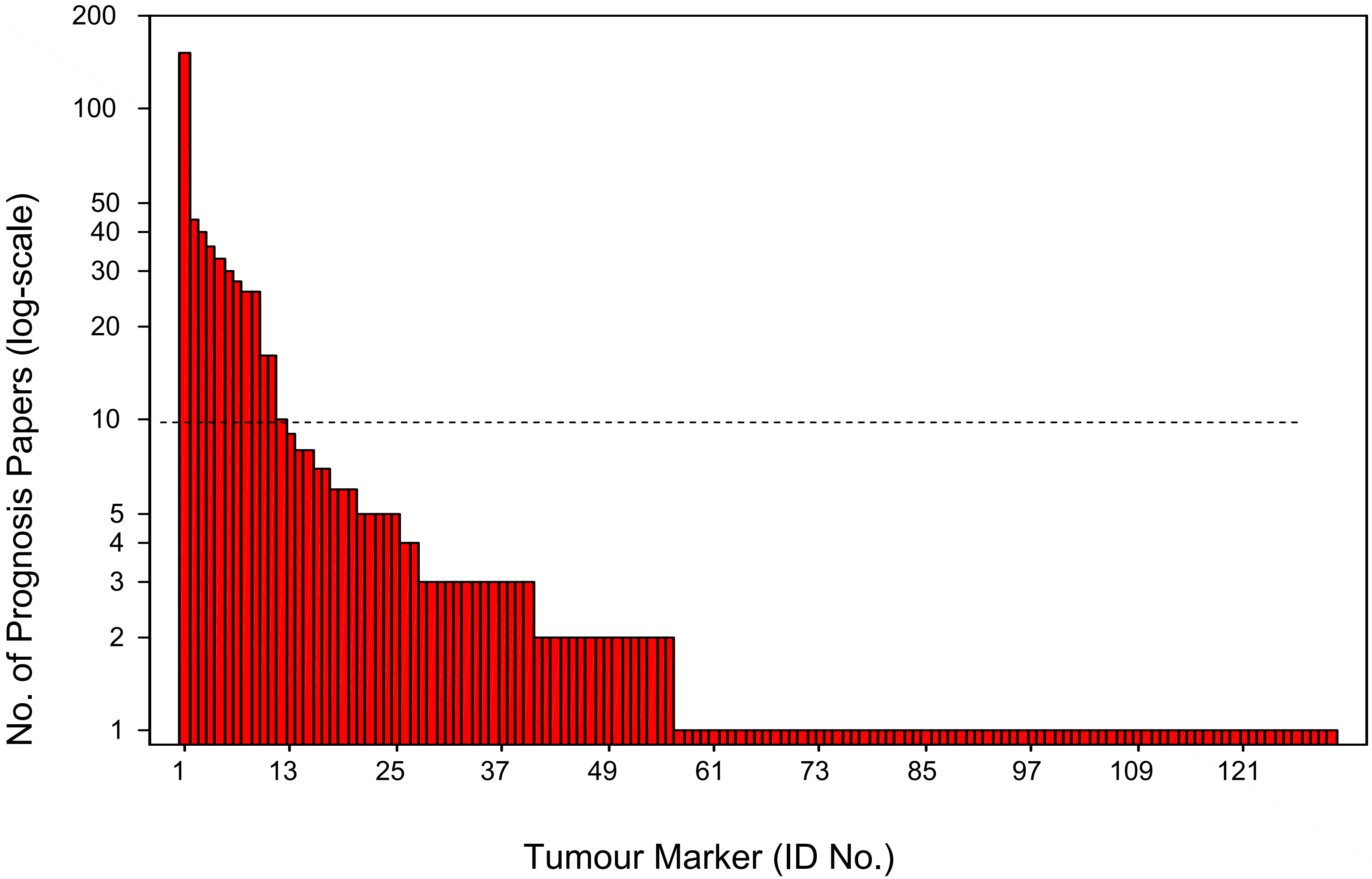
- *Search strategy* → *Medline/Embase/Cancerlit (1966 to 2000)*
  → **3415 papers identified**

- *Inclusion/Exclusion* → **260 prognosis papers**

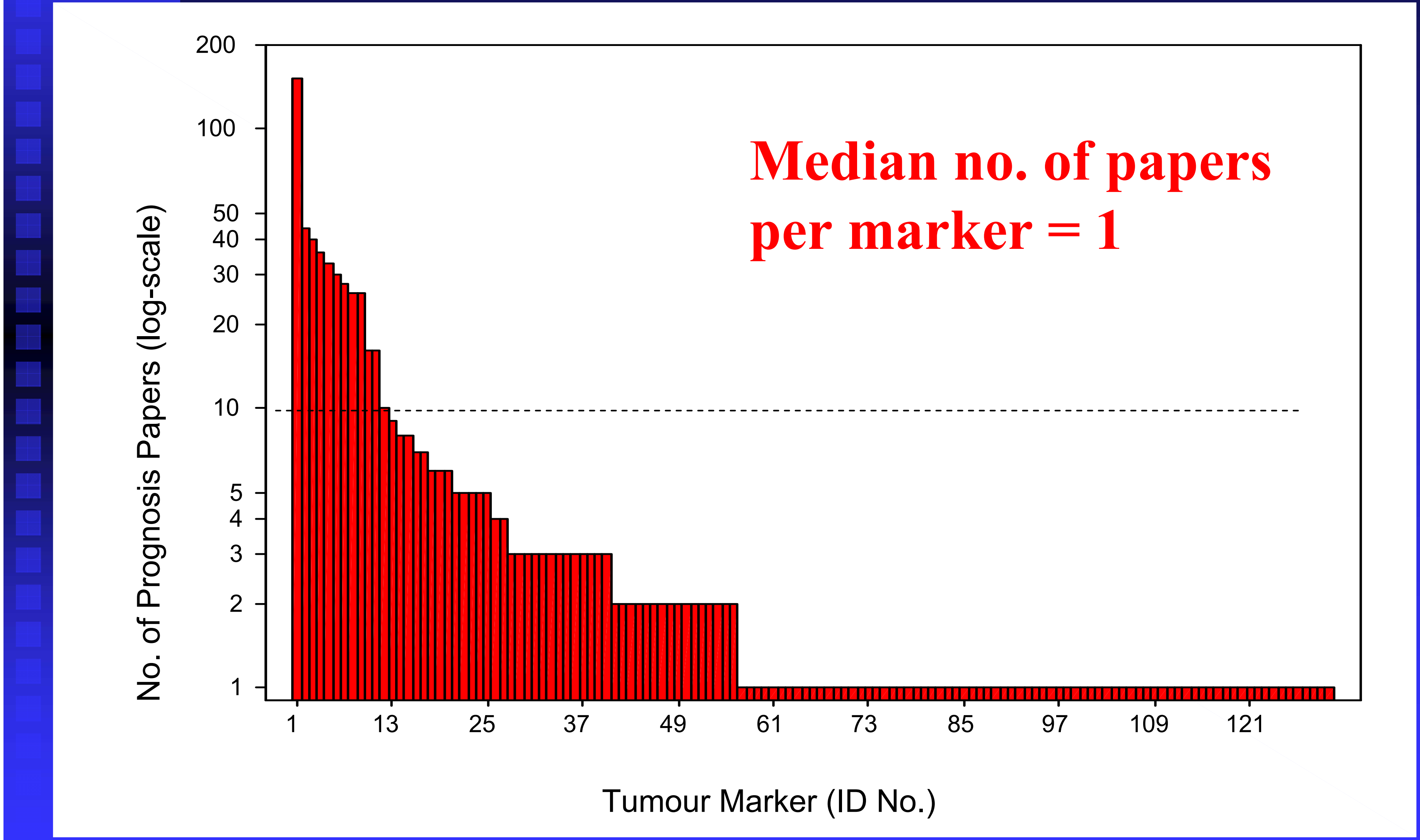
131 different prognostic markers studied in the 260 papers identified

This emphasises the need for evidence-based research

13 markers most commonly reported were selected for further study



13 markers most commonly reported were selected for further study



Median no. of papers per marker = 1

What summary statistic to extract?

Need a statistic that compares time to death and/or recurrence of disease in different risk groups

- ***Hazard ratio*** desirable because:

  1) *Relative risk* for survival data

  2) estimate (with standard error) of the difference in outcome between groups of patients defined by the marker

  3) Takes into account the whole follow-up period, not just one specific time-point

  4) Incorporates those patients censored (lost to follow-up)

  5) log(HR) approx Normal, aiding meta-analysis models

Extracting the Hazard Ratio .

Unadjusted or adjusted Hazard Ratio?

- **Unadjusted** hazard ratios were sought from each paper
- If not possible, adjusted results were then sought

Extract the estimates required from each study

- Papers commonly reported > 1 result,

  e.g. for **more than one marker**

  for **both** *overall* (OS) and *disease-free survival* (DFS)
- Estimates for both OS and DFS desired

What is the overall evidence for each of the 13 markers?

→ **260 different published prognosis studies**

→ **575 results or IPD from which an OS or DFS hazard ratio desired for one of the 13 markers**

Extracting the Hazard Ratio and Variance

1) Easy if they have presented the hazard ratio & variance directly:

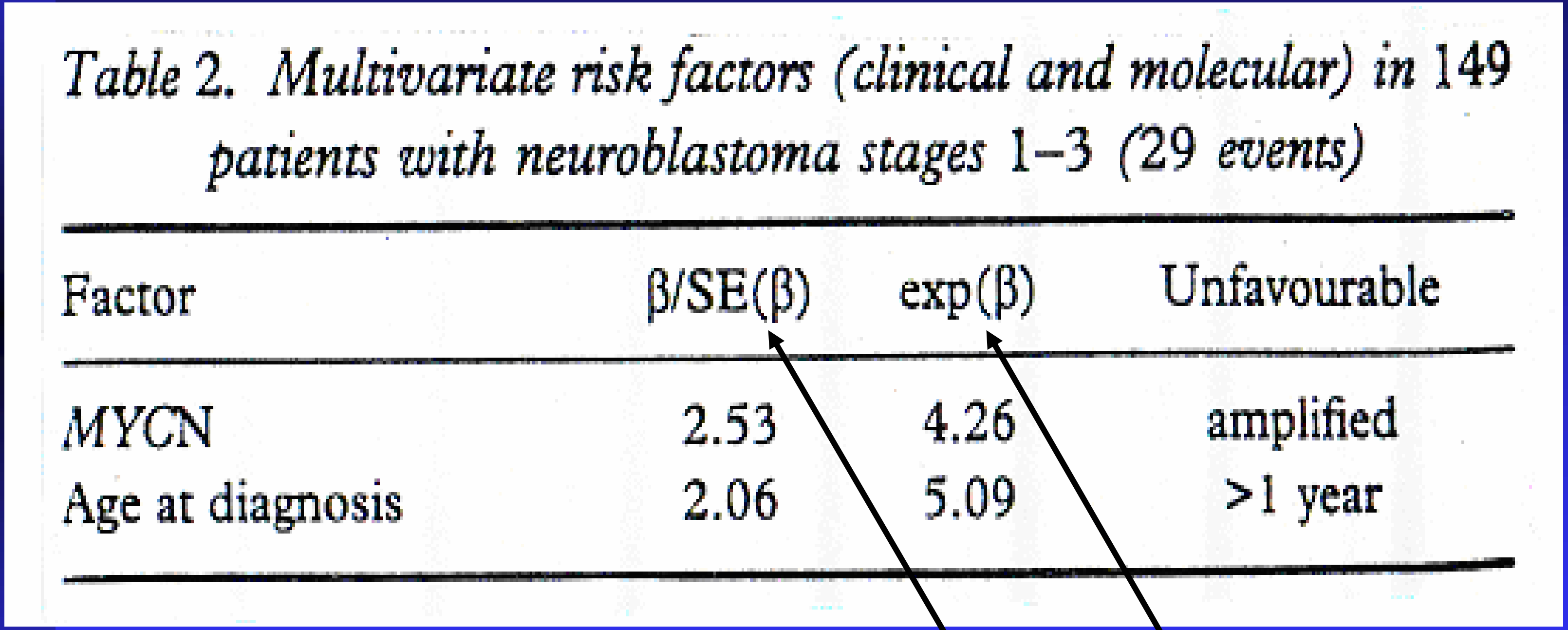


Table 2. Multivariate risk factors (clinical and molecular) in 149 patients with neuroblastoma stages 1–3 (29 events)

Factor	β/SE(β)	exp(β)	Unfavourable
MYCN	2.53	4.26	amplified
Age at diagnosis	2.06	5.09	>1 year

3 out of 575 (0.52%)

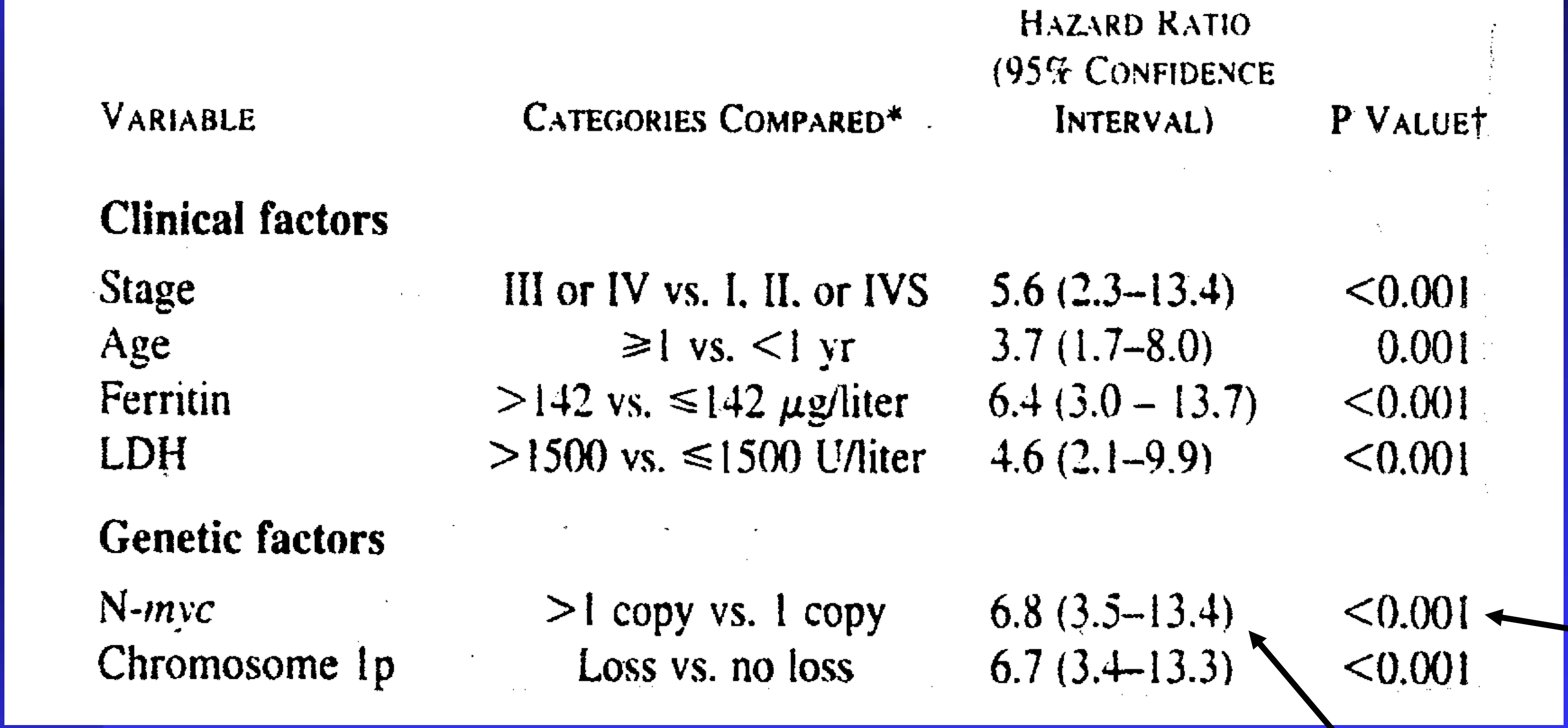
(all from just 1 paper out of 260)

hazard ratio (HR)

variance of log(HR)

2) **Indirect** estimation needed (**Parmar et al, 1998**):

(i) *Hazard ratio & CI,* *or* *(ii)* *Hazard ratio & p-value*



Variable	Categories Compared*	Hazard Ratio (95% Confidence Interval)	P Value†
Clinical factors			
Stage	III or IV vs. I, II, or IVS	5.6 (2.3–13.4)	<0.001
Age	≥1 vs. <1 yr	3.7 (1.7–8.0)	0.001
Ferritin	>142 vs. ≤142 μg/liter	6.4 (3.0 – 13.7)	<0.001
LDH	>1500 vs. ≤1500 U/liter	4.6 (2.1–9.9)	<0.001
Genetic factors			
N-*myc*	>1 copy vs. 1 copy	6.8 (3.5–13.4)	<0.001
Chromosome 1p	Loss vs. no loss	6.7 (3.4–13.3)	<0.001

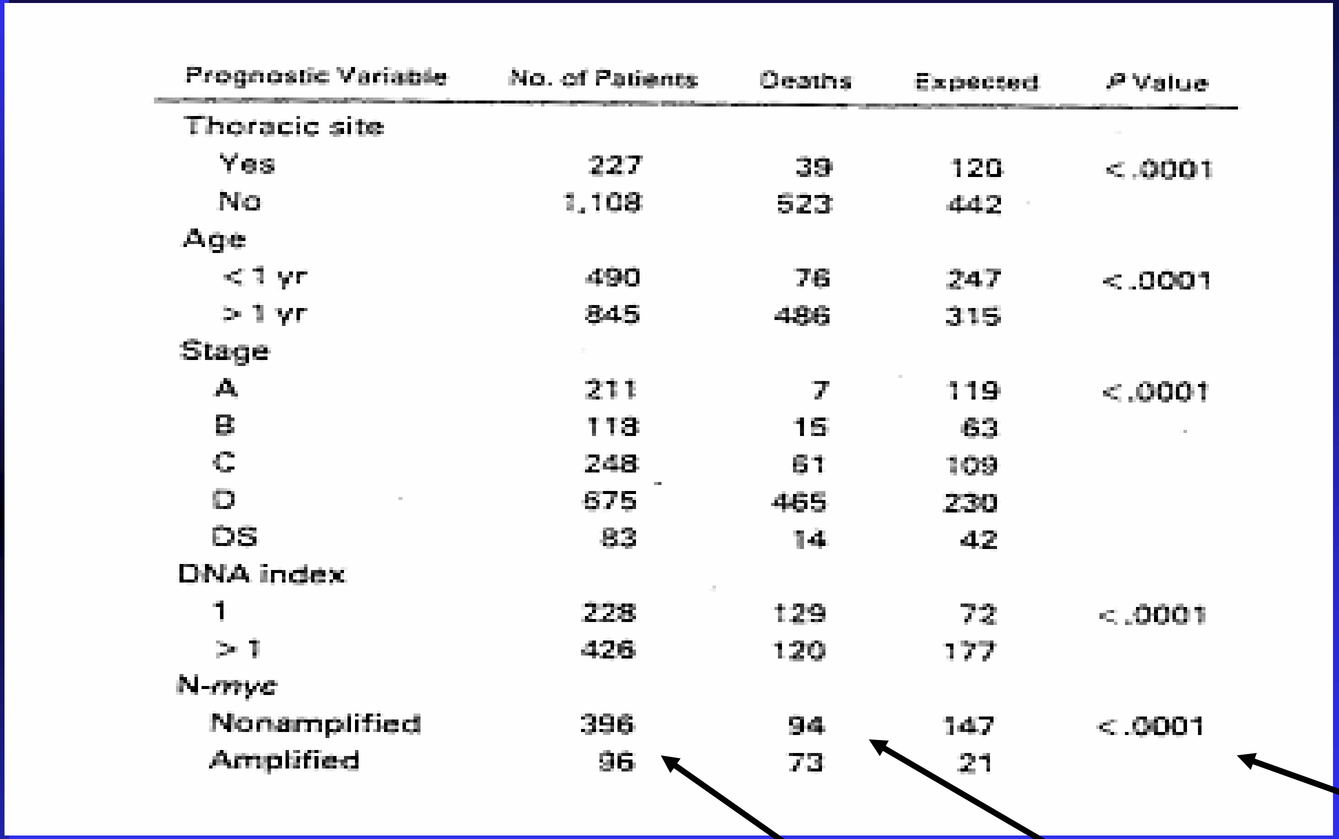
P-value

HR and confidence interval

52 out of remaining 572 (9.0%)

Cumulative: 9.6% of the 575

(3) *Use P-value, group sizes and group events*



Prognostic Variable	No. of Patients	Deaths	Expected	*P* Value
Thoracic site				
Yes	227	39	120	<.0001
No	1,108	523	442	
Age				
< 1 yr	490	76	247	<.0001
> 1 yr	845	486	315	
Stage				
A	211	7	119	<.0001
B	118	15	63	
C	248	61	109	
D	675	465	230	
DS	83	14	42	
DNA index				
1	228	129	72	<.0001
> 1	426	120	177	
N-*myc*				
Nonamplified	396	94	147	<.0001
Amplified	96	73	21	

P-value

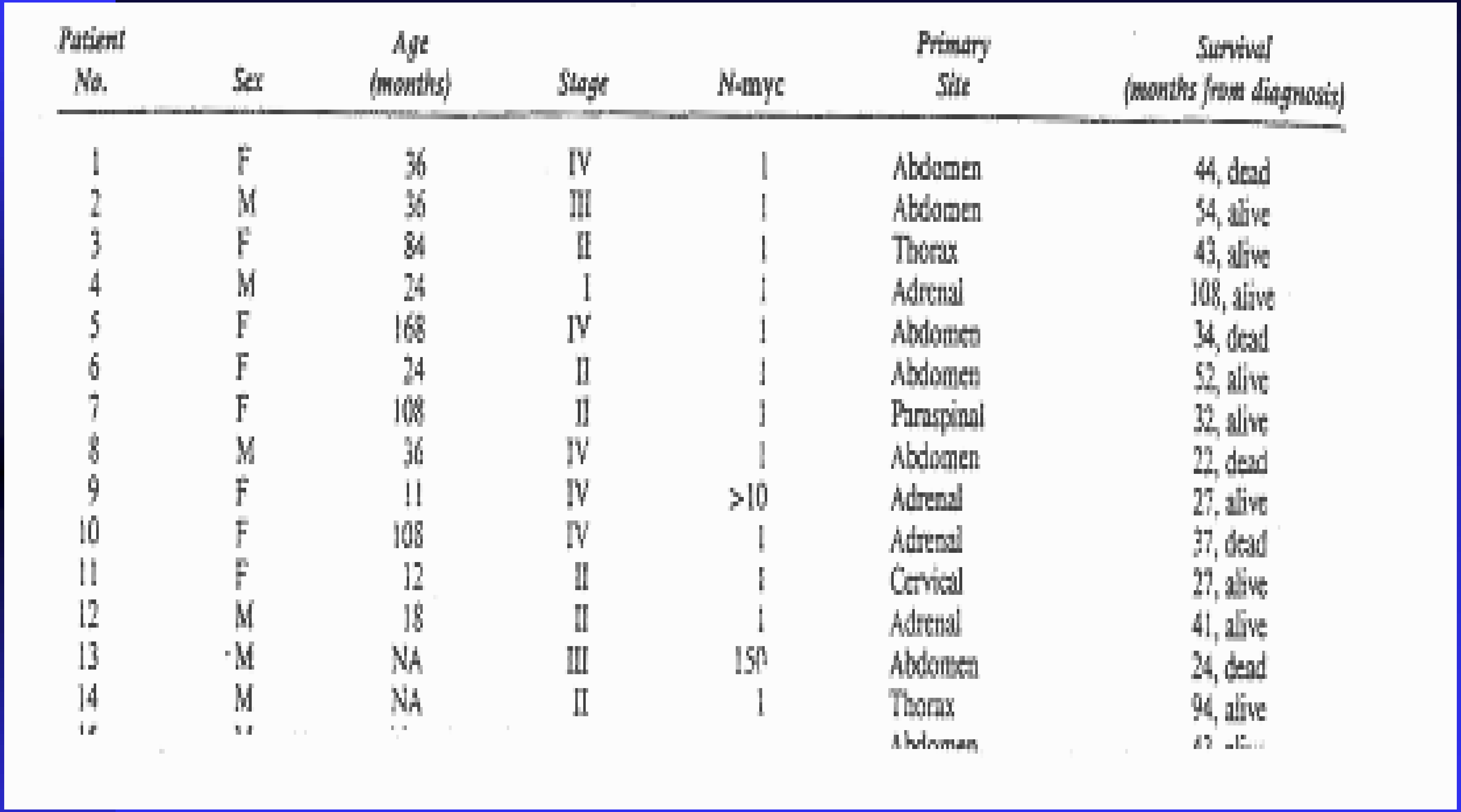
No. events

No. patients

104 out of remaining 520 (20%)

Cumulative: 27.7% of the 575

(4) *Use **Individual Patient Data** to fit a Cox regression model*



Patient No.	Sex	Age (months)	Stage	N-myc	Primary Site	Survival (months from diagnosis)
1	F	36	IV	1	Abdomen	44, dead
2	M	36	III	1	Abdomen	54, alive
3	F	84	II	1	Thorax	43, alive
4	M	24	I	1	Adrenal	108, alive
5	F	168	IV	1	Abdomen	34, dead
6	F	24	II	1	Abdomen	52, alive
7	F	108	II	1	Paraspinal	32, alive
8	M	36	IV	1	Abdomen	22, dead
9	F	11	IV	>10	Adrenal	27, alive
10	F	108	IV	1	Adrenal	37, dead
11	F	12	II	1	Cervical	27, alive
12	M	18	II	1	Adrenal	41, alive
13	M	NA	III	150	Abdomen	24, dead
14	M	NA	II	1	Thorax	94, alive
[illegible]	[illegible]	[illegible]			Abdomen	[illegible]

41 of remaining 416 (9.9%)

Cumulative: 35.3% of the 575

(5) *Survival curve extraction*

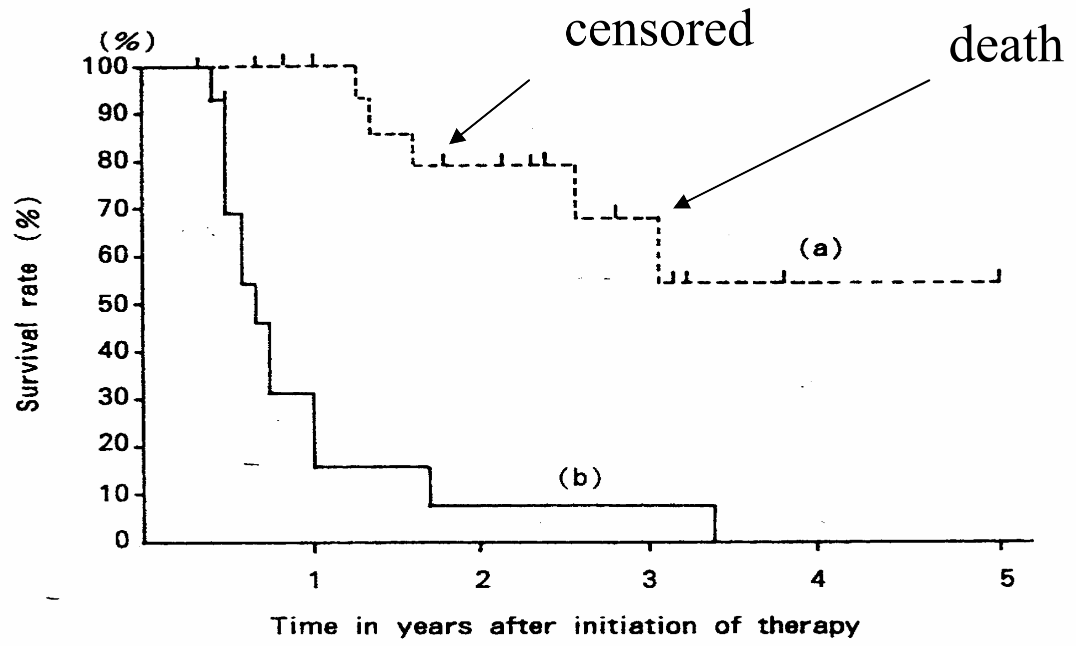


Fig. 1. N-myc amplification and cumulative survival curves of patients with stages III and IV neuroblastoma. *(a)* L-group: N-myc; 1-10 copies, $n = 18$; *(b)* H-group: N-myc; > 10 copies, $n = 14$; $p < 0.001$.

4 of remaining 375 (1.1%)

Cumulative: 35.5% of 575

Summary of the Overall Number Obtained

- **204 estimates obtained out of 575 desired**
- For the other 371 (**64.5%**) we could not extract a hazard ratio by any of the above methods

Could we have done anything else?

- e.g. use % survival at n years for:
  (i) estimating hazard ratio, or
  (ii) as the statistic in the meta-analysis

- The benefit of this is marginal
- % survival is equally poorly reported

e.g. in 26 prognosis papers for marker LDH:

- 12 gave actuarial estimates of % survival
- only 6 of these gave a confidence interval or standard error
- 4 different time-points used: 2,3,4,5 years

Problem for Meta-analysis No. 1

Poor Reporting of Primary Studies

- Prevents a reliable meta-analysis
- Can not include all the evidence
- Only 35.5% of estimates obtained
- Two thirds of the evidence not available
- May introduce bias

What about more recent studies?

- Reporting has improved
- e.g. the 26 papers giving a hazard ratio published > 1990
- Yet, still represents only 17% of the total literature since 1990 assessed

Key Reporting Problems

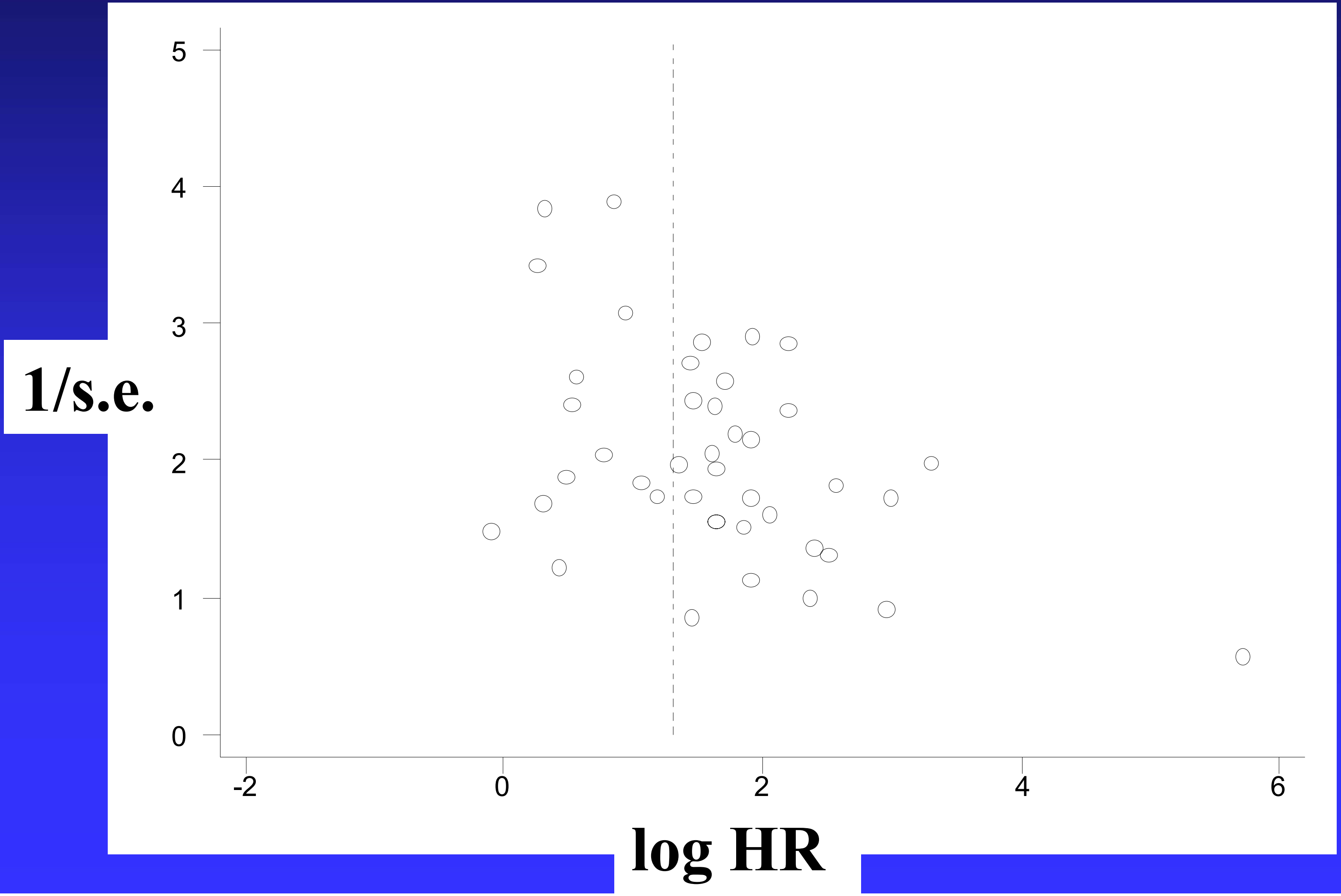
- No appropriate statistical analysis performed or reported
- Hazard ratio not calculated or not reported
- Just p-value provided and not confidence intervals
- Inexact p-values provided, e.g. p<0.05 or ‘significant’
- Group numbers and group events not given

Why is this happening?

- Lack of statisticians involved
- Lack of statistical knowledge, understanding and ability
- Lack of guidelines on how to do things better
- Unaware of why improved reporting is needed
- Focus on obtaining publications from primary studies
- No understanding of evidence-based research
- Biased and selective reporting of results?

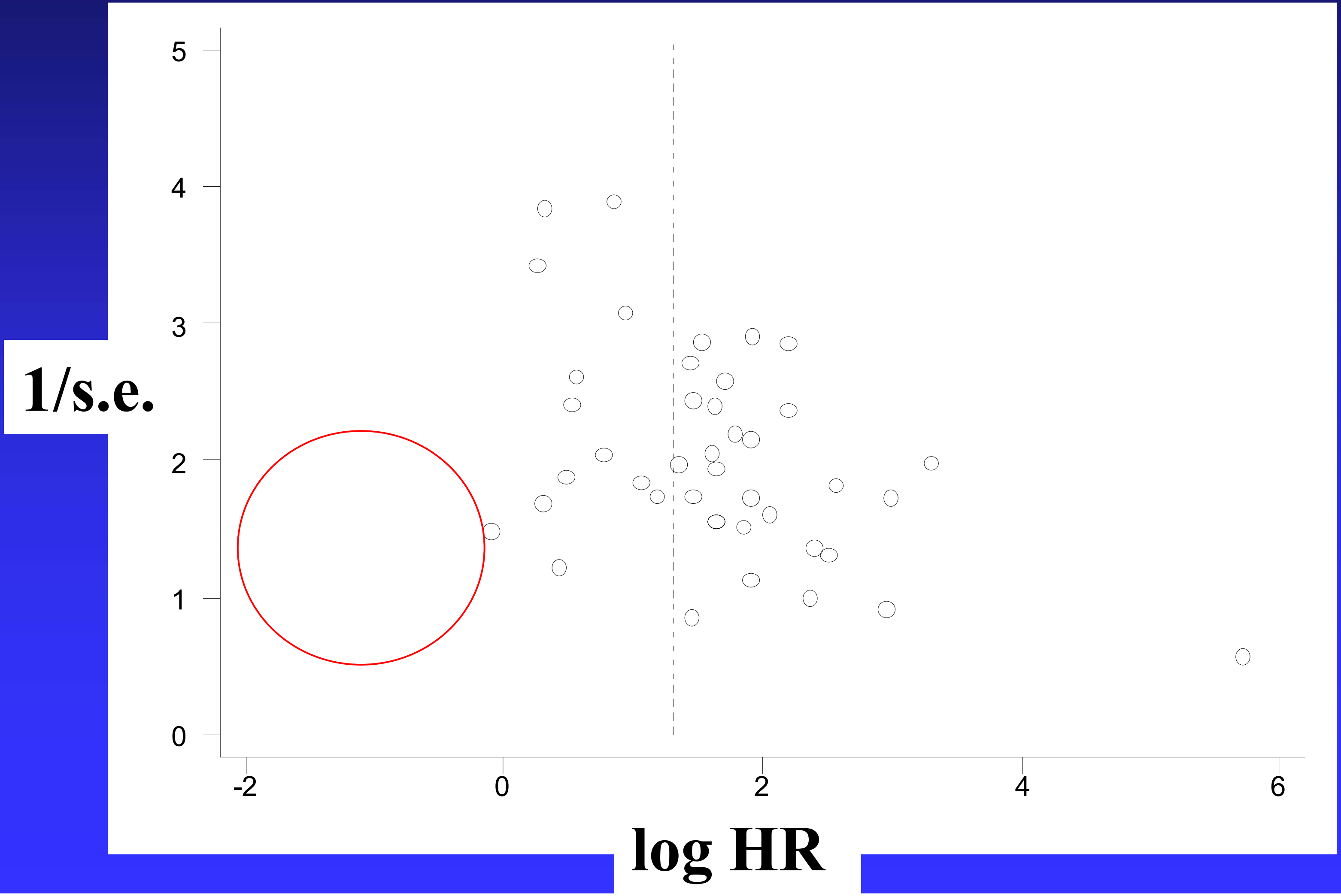
Evidence of small study effects (publication bias?)

- marker MYCN & disease-free survival
- hazard ratio & s.e. obtained for 42 studies



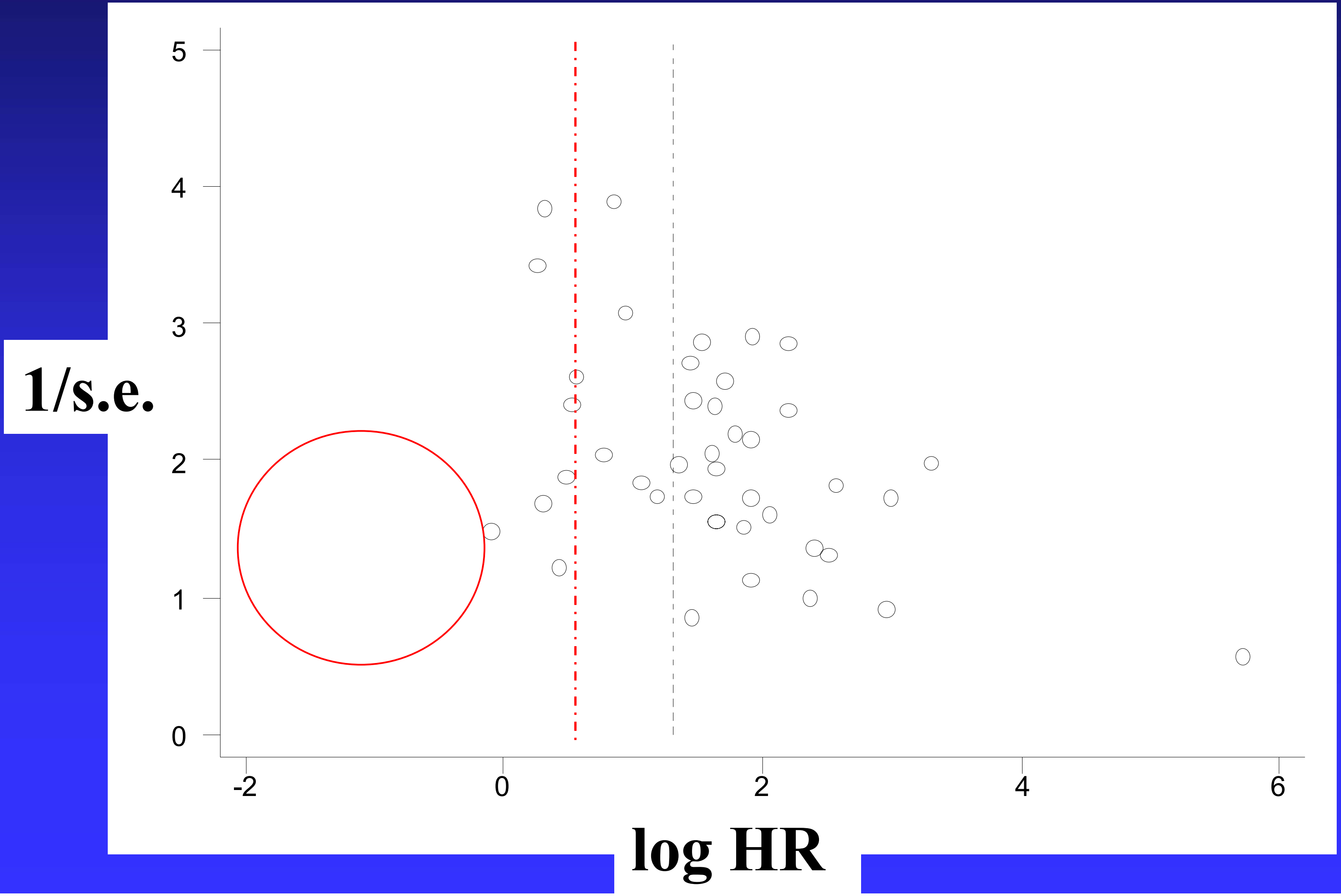
Evidence of small study effects (publication bias?)

- marker MYCN & disease-free survival
- hazard ratio & s.e. obtained for 42 studies



Evidence of small study effects (publication bias?)

- marker MYCN & disease-free survival
- hazard ratio & s.e. obtained for 42 studies



Other evidence of reporting & bias problems

- Kyzas et al. (2005) review 331 cancer prognostic studies

  - conclude that the reporting of study design and assay information was often suboptimal

- Kyzas et al. (2007) review 1915 prognostic marker articles

  - nearly all articles present significant findings

  < 1.5% were fully 'negative' in that they did not present statistically significant prognostic results and did not elaborate on non-significant trends.

Other evidence of reporting & bias problems

- A systematic review of studies of Bcl2 in non-small cell lung cancer (Martin et al., 2003)

  Small studies

  - all show a statistically significant relationship between Bcl2 and death

  Large studies

  - all 3 are non-significant & show a smaller effect

Other evidence of reporting & bias problems

- Simon (2001) comments that the prognostic literature:

  "is probably cluttered with false-positive studies that would not have been submitted or published if the results had come out differently."

- Rifai et al. (2008) believe that is time to take action against reporting biases in prognostic studies.

- Guidelines have been proposed ...

REMARK guidelines (McShane et al., 2005)

- Aim to improve reporting standards

- Suggest the key information to be reported from a prognostic marker study

- Considers the whole study process

  from pre-defined hypotheses and patients included
  ... to the statistical methods used and results identified
  ... to the study limitations and implications for practice

- Journal editors are encouraged to enforce REMARK

Other guidelines for improved reporting

- Riley et al. (2003); Altman et al. (1995)

  - how aggregate data should be reported
  - effect estimates & confidence intervals
  - clear presentation of survival curves
  - details of adjustment factors

An Example of Better Reporting

	Variable	Alive	Dead	Total	Hazard Ratio (HR)	95% CI	P-value
TH	-ve	4	12	16	2.40	1.19 to 4.84	0.014
	+ve	4	29	33			
NSE status	<100	2	9	11	1.45	0.64 to 3.28	0.38
	>=100	4	16	20			
	Missing	2	16	18			
LDH Status	<1500	4	9	13	11.11	3.30 to 37.4	0.0001
	>=1500	0	11	11			
	Missing	3	22	25			
Ferritin Status	<150	2	4	6	1.92	0.65 to 5.66	0.24
	>=150	3	21	24			
	Missing	3	16	19			
Age	1-2	3	10	13			0.003
	2-3	0	12	12	4.09	1.58 to 10.62	
	3-5	4	12	16	0.80	0.34 to 1.87	
	>5	1	7	8	0.88	0.33 to 2.32	
NMYC	-ve	5	21	26	1.38	0.65 to 2.93	0.41
	+ve	1	10	11			
	Missing	2	10	12			
Overall		8	41	49			

Other guidelines for improved reporting

- Riley et al. (2003); Altman et al. (1995)

  - how aggregate data should be reported
  - effect estimates & confidence intervals
  - clear presentation of survival curves
  - details of adjustment factors

- Burton et al. (2004)

  - encourage clearer reporting of missing data

- Numerous authors encourage availability of IPD

Problem for Meta-analysis No. 2

Heterogeneity of clinical and statistical factors

In the neuroblastoma review, of the 204 estimates obtained there was great variability in:

CLINICAL & REPORTING FACTORS: e.g.

- Cut-off level used to dichotomise the continuous markers
- Method of measuring the marker

- Stage of disease
- Age of Patients

- Type of treatment received
- Outcome – overall or disease-free survival

Problem for Meta-analysis No. 2

Heterogeneity of clinical and statistical factors

In the neuroblastoma review, of the 204 estimates obtained there was great variability in:

STATISTICAL FACTORS:

- Type of estimate,
  e.g. unadjusted and adjusted; indirect and direct
- Adjustment factors
- Analysis method

DESIGN FACTORS

- Study design (e.g. Method of marker measurement)
- Purpose of the study
- Study quality

Example of heterogeneity in the 94 estimates obtained for marker MYCN

		n			n
Outcome	DFS	46	**Cut-off Point**	1 copy	23
	OS	48		2 copies	1
				3 copies	17
Result Type	unadjusted	77		4 copies	5
	adjusted	17		5 copies	2
				10 copies	18
Stage groups	all	68		Mean gene expression	2
	1	2		Positive vs negative protein (or staining vs no staining)	9
	3	2			
	4	4		unknown	17
	1, 2, 3	3			
	1, 2, 3, 4	5	**Age groups**	all	78
	2, 3, 4, 4S	2		< 1year	2
	3, 4	3		> 1 year	5
	unknown	5		unknown	9

The Dilemma for meta-analysis

- This heterogeneity exists in addition to the incomplete set of evidence from the poor reporting

- Is it right to pool the estimates available?

- Clinical and statistical interpretation of meta-analysis results difficult

- **Could we make strong clinical recommendations?**

- e.g. clear results for specific stages of disease?
- e.g. marker X is better than marker Y?
- e.g. marker X should be used in addition to marker Y?

- Compounded by issue of publication/reporting bias

Overcoming the Problem of Heterogeneity

- Collaboration of research groups required
- Seek consistency in cut-offs, adjustment factors, outcomes, analysis, measurement methods etc.
- Multi-disciplinary teams
- Improve study design standards (Altman and Lyman, 1998) – e.g. protocol driven
- Design large prospective studies to answer prespecified questions of clinical interest
- Promote better reporting

Overcoming the Problem of Heterogeneity

- Large, prospective multi-centre studies
- Facilitate access to tumour banks, containing detailed patient-level information
- Collaborate across research groups & pool IPD
(e.g. breast cancer - Look et al., 2002)
- Prospectively planned pooled analyses
- seek common aims
- agree common design and clinical factors
- agree to pool IPD at the end

The Benefit of Having IPD From Each Study

- IPD would limit poor reporting by allowing:
  - data checking
  - consistent statistical analysis in each study
  - model assumptions to be verified
  - estimates of interest to be calculated
  - proper handling of continuous variables
- IPD would limit heterogeneity in:
  - type of estimates (adjusted/unadjusted)
  - outcome
  - adjustment factors
  - cut-off level (use continuous level?)
- IPD facilitates
  - analysis of subgroups (e.g. age < 1)
  - analysis of combinations of markers

What to include in the IPD?

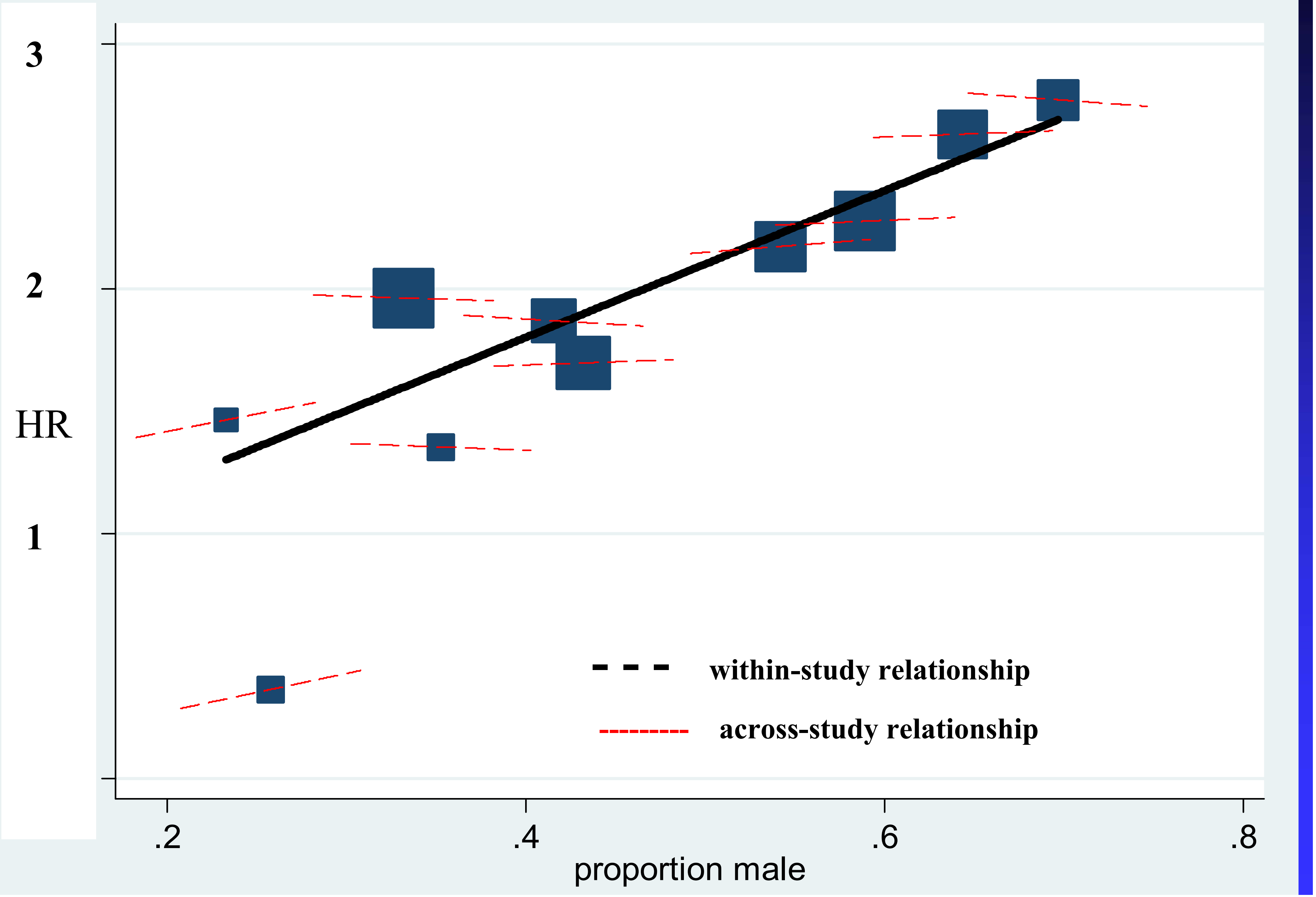
For all markers considered (not just those ‘significant’), include:

- Relevant patient details (e.g. age, stage)
- exact initial marker level and how marker was measured
- time of disease recurrence (if appropriate)
- follow-up time
- final disease status
- important adjustment factors
- treatment received

An example IPD

	Marker levels				Adjustment factors			Survival and disease status		
no.	TH	LDH	MYCN	…	Age	Stage	…	Time of recurrence	Final survival time	Final disease status
1	Pos	200	5		3 yrs	1		-	150 days	ALIVE
2	Neg	350	3		2 yrs	4		330 days	390 days	DEAD
3	Neg	120	1		2 yrs	3		230 days	250 days	ALIVE with disease
4	Neg	320	1		6 yrs	4		27 days	48 days	DEAD
…	…	…	…		…	…		…	…	…

Difference between within-trial and across-trial relationships



IPD – am I being realistic?

- Researchers **protective** over their own data
- Worried about **Data Protection Act** – no need to put in ID number.
- **Cost, time** – when does it become worthwhile?
- To identify the best prognostic markers we need to be prepared to collaborate and share data.
- Try to make IPD available - in paper, on Web, on request
- Be involved in prospectively planned pooled analyses

Generalisations to other diseases

- Altman (1995) shows a general standard of poor reporting in survival studies

- Other systematic reviews of prognostic markers limited
  e.g. Lung cancer (Brundage et al., 2002)
  - *median no. of papers per marker = 1*

  Brain damage (Zanbergen et al., 2001)
  – *small samples & different laboratory techniques*

  Prostate cancer (Parker et al, 2001)
  – *incomplete & heterogeneous nature of reports*

- Increasing evidence of reporting biases (e.g. Kyzas work)

- Lack of consensus regarding design standards

Generalisations to other diseases

- Schmitz-Dräger et al. (2000) review 43 trials regarding p53 immunohistochemistry as a prognostic marker in bladder cancer

- Conclusion:

“From this analysis it becomes evident that further retrospective investigations will not contribute to the solution of the problem and thus are obsolete.

There is an obvious need for standardization of the assay procedure and the assessment of the specimens as well as for the initiation of a prospective multi-centre trial to provide definite answers.”

Reasons to be optimistic

- IPD can be obtained, although may be a long process (Altman et al., 2006)
- Meta-analyses have been facilitated when IPD available

  e.g. in determining a consistent cut-off level

  (Sakamoto et al., 1996; Look et al., 2003)

- Awareness of reporting biases (e.g. Kyzas work)
- Design guidelines & identification of 'Phases' of prognosis research

  (e.g. Altman and Lyman, 1998; Hayden et al., 2008)

- Reporting guidelines (e.g. REMARK)

Reasons to be optimistic

- The initiation of tumour banks
- Hayes et al. (2008) state that the exciting potential of prognostic markers highlights the *"importance of prospective collection, processing, and storage of biospecimens"*
- e.g. Goebell et al. (2004): establishing a multi-institutional bladder cancer database & virtual tumour bank to evaluate the prognostic significance of potential markers.
- Many others too; e.g. Confederation of Cancer Biobanks

Reasons to be optimistic

- This meeting!
- Cochrane Prognosis Methods Group

- Aims to facilitate evidence-based prognosis research

- Improve design, quality & reporting of primary studies

- Facilitate systematic reviews & meta-analysis in long-run

- Bring together prognosis researchers

- Please join!

Summary

- Evidence-based use of prognostic markers essential

- Systematic reviews & meta-analysis limited

  - Poor reporting
  - Publication bias & selective reporting
  - Small, poorly designed primary studies
  - Statistical, clinical & methodological heterogeneity

- Guidelines for improvement
- Availability of IPD necessary
- Work together – multiple disciplines (involve editors)
- Multi-centre studies
- Prospective meta-analysis

e-mail: richard.riley@liv.ac.uk

Select References

- Altman DG: Systematic reviews of evaluations of prognostic variables. BMJ 323:224-8, 2001
- Kyzas PA, et al: Almost all articles on cancer prognostic markers report statistically significant results. Eur J Cancer 43:2559-79, 2007
- Kyzas PA, et al: Selective reporting biases in cancer prognostic factor studies. J Natl Cancer Inst 97:1043-55, 2005
- McShane LM, et al: REporting recommendations for tumor MARKer prognostic studies (REMARK). J Natl Cancer Inst 97:1180-4, 2005
- Riley RD, et al: Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. Br J Cancer 88:1191-8, 2003
- Riley RD, et al: Prognosis research: toward evidence-based results and a Cochrane methods group. J Clin Epidemiol 60:863-5; 2007
- Sauerbrei W, et al: Evidence-based assessment and application of prognostic markers: the long way from single studies to meta-analysis. Communications in Statistics 35:1333-1342, 2006