

Empirical Studies of Inspection and Test Data at Ericsson

Amarjit Singh Marjara, Cap Gemini Norway AS

Reidar Conradi, NTNU;

Børge Skåtevik, STC

NASA SEL Workshop, 1-2 Dec 1999

(rev. 8 Dec. 99)

GSFC, Washington

Agenda

- Background
- The inspection method
- Data
- Observations/questions
- Results
- Conclusions
- Recommendations

Purpose of the studies

- **H₁**: To investigate if there is a correlation between defects found during inspection/test and the complexity.
- **H₂**: To investigate if there is a correlation between the number of defects found in field-use and the complexity and the modification rate of a module.
- **H₃**: To investigate if there is a correlation between defects rates across phases and deliveries for individual documents/modules.
- A diploma work at NTNU Sept.-Dec. 1997, and diploma work at NTNU Oct'98-Feb'99, against Ericsson AS, Norway.

Background

- Data are collected at Ericsson AS, AXE-division, Oslo.
- Every development document (design, code,..) is inspected.
- Using Gilb method; an extension of Fagan's.
- Data for many projects are analysed.
- The analysed data originates from design, unit test, function test and system test. Code is not inspected in this manner.

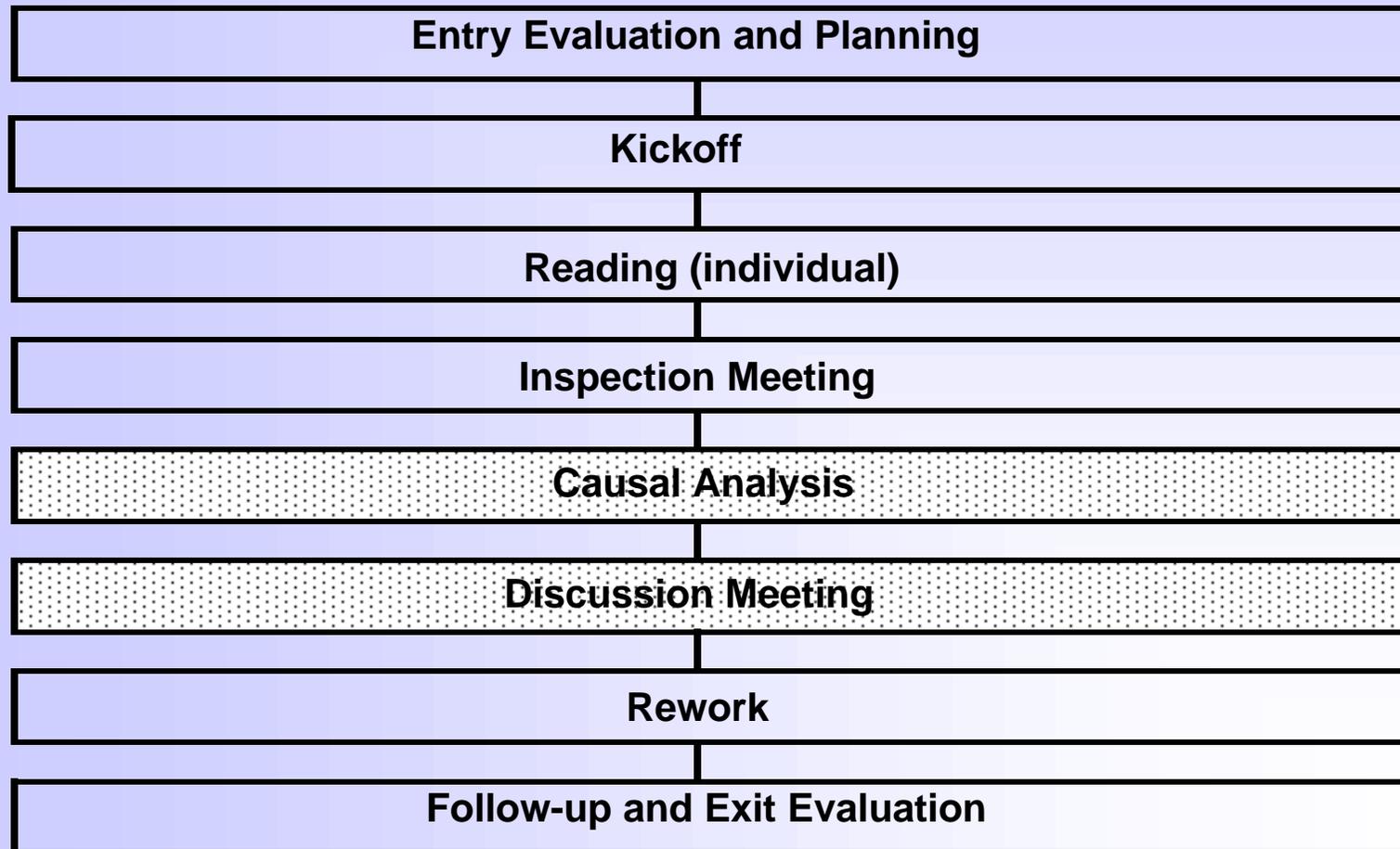
Background 2(3)

- The paper is divided in two studies:
 - Study 1:
 - Data from one project of 20.000 person-hours.
 - It includes design inspections, desk check (code review & unit test), function test, and partly system test and field-use.
 - The initial phases, such as prestudy and system study, are excluded from 20.000 ph.

Background 3(3)

- Study 2, done later:
 - A study of 6 projects ==> 100.000 ph. (including the one in Study 1).
 - Includes design inspections, code review, and unit test -- not function test etc.
 - The initial phases, such as prestudy and system study, excluded from 100.000 ph.

The inspection method 1(2)



The inspection method - 2(2)

- Provide special training for the moderators.
- Inspection meeting max two hours.
- Follow the recommended, “optimal” reading rates for the actual document type (18 such).
- Do not cover too much complex material in a single review.
- Invite the most competent inspectors to the meeting.
- Avoid personal criticisms.
- Postpone long discussions till end of the meeting.

Inspection Data

- **Block** - name of the block (module, unit).
- **Document type** - the type of document which is inspected.
- **Effort saved/lost** - estimating whether effort has been saved or not.
- **Number of persons** - doing reading and participating in the inspection meeting.
- **Planning** - effort spent on planning the inspection.
- **Kickoff meeting** - effort spent on introducing the document to the participants.
- **Reading** - total effort spent on individual preparation.
- **Inspection meeting** - group effort.
- **Rework, follow up** - defect fixing effort.
- **Defects found during reading** : classified (Super Major or Major)
- **Defects found in inspection meeting, pages read, pages per meeting**
 - **Defect classification: also just Super Major or Major**
- **====> All summarized in an inspection survey, stored in a DB**

Collecting Data - testing

■ Unit test

- No. of defects found and effort spent
- Data available per module (=unit)

■ Function test, system test, field-use

- Cause
- Priority - seriousness of the defect
- Data available per module
- The effort spent on system test and field-uses not available (final integration in Stockholm).

Results

Study 1 - effectiveness of inspections and testing

■ Results from Study 1:

| Activity | Defects [#] | [%] |
|--------------------------------------|-------------|--------------|
| Individual reading (design) | 928 | 61.8 |
| Inspection meeting (design) | 29 | 1.9 |
| Desk check (code review + unit test) | 404 | 26.9 |
| Function test | 89 | 5.9 |
| System test | 17 | 1.1 |
| Field use | 35 | 2.3 |
| Total | 1502 | 100.0 |

Note: 9.3% of defects in three latter phases.

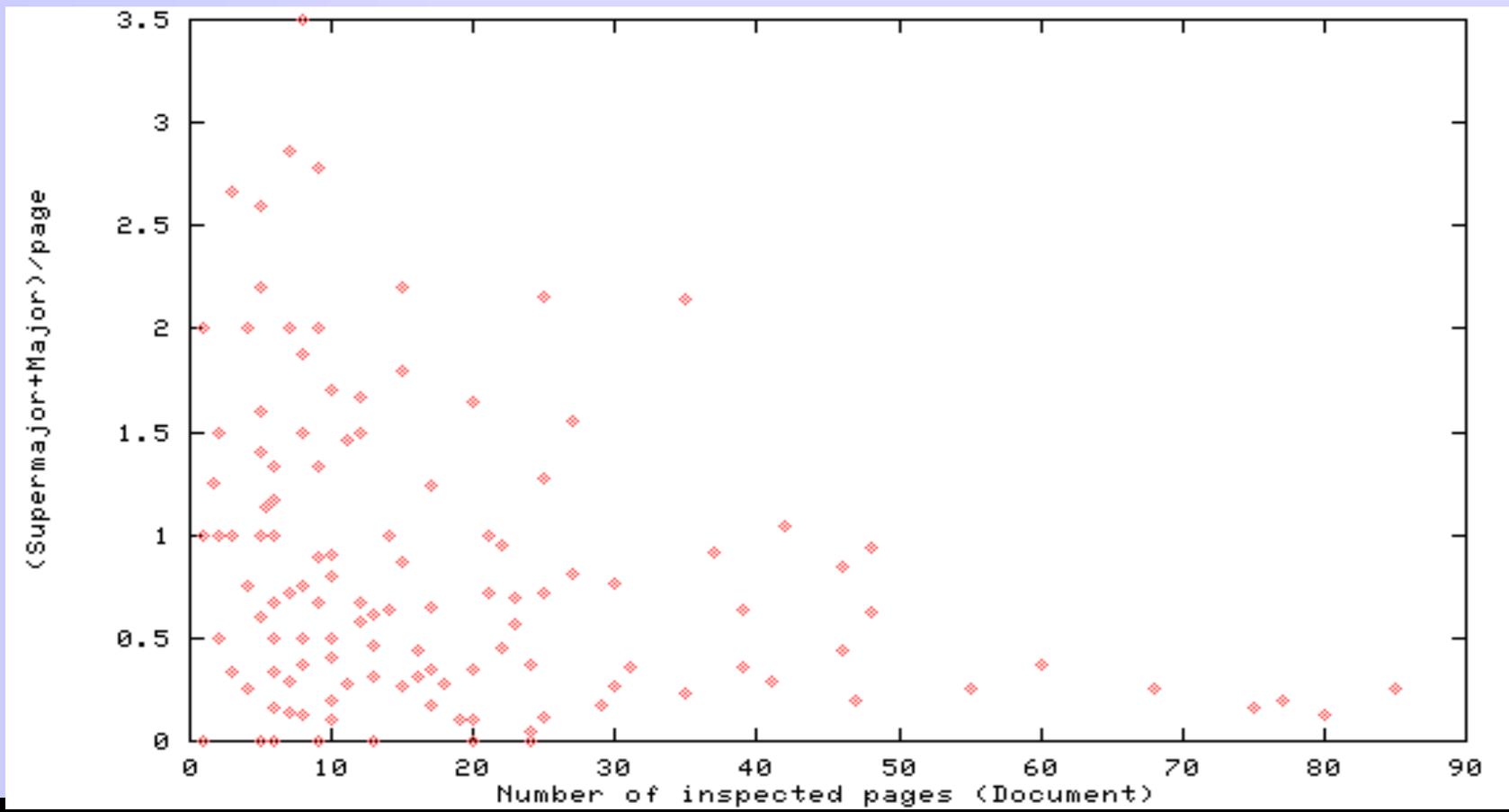
Results

Study 1- Cost-effectiveness of inspections and testing

| Activity | Defects [#] | Effort [h] | Of total [%] | Effort spent to find one defect [h:m] | Estimated saved effort by early defect removal [h] |
|-----------------------------|-------------|------------|--------------|---------------------------------------|--|
| Inspection reading (design) | 928 | 786.8 | 61.78 | 00:51 | 8200 |
| Inspection meeting (design) | 29 | 375.7 | 1.92 | 12:57 | |
| Desk check | 404 | 1257.0 | 26.91 | 03:07 | - |
| Function test | 89 | 7000.0 | 5.92 | 78:39 | - |
| System test | 17 | - | 1.13 | - | - |
| Field use | 35 | - | 2.33 | - | - |

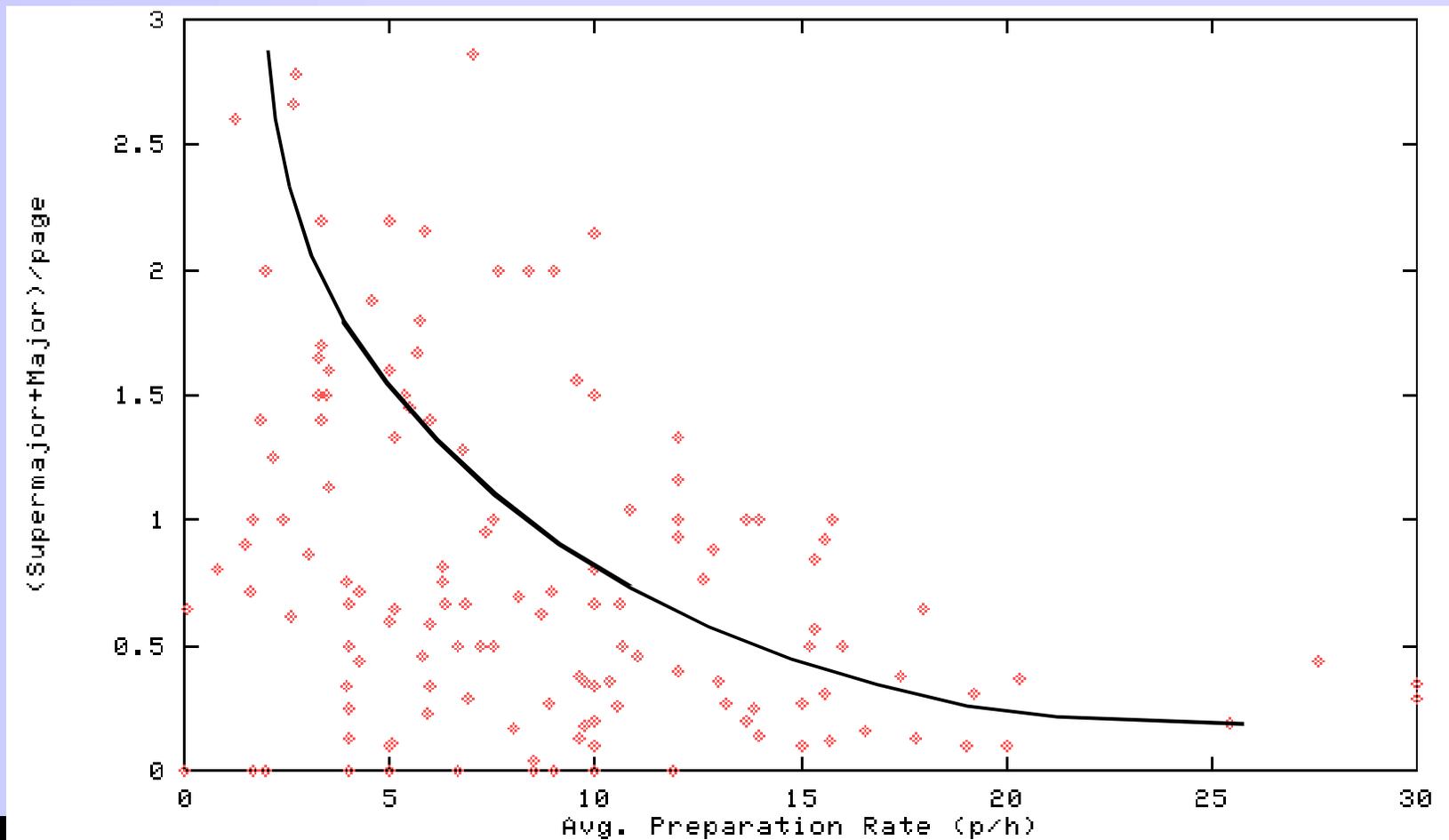
Results

Study 1 - Number of pages inspected and defect detection rate



Results

Study 1- Reading rate and defect detection rate



Results

Study 1- field-use defects and defects found during inspections, per module

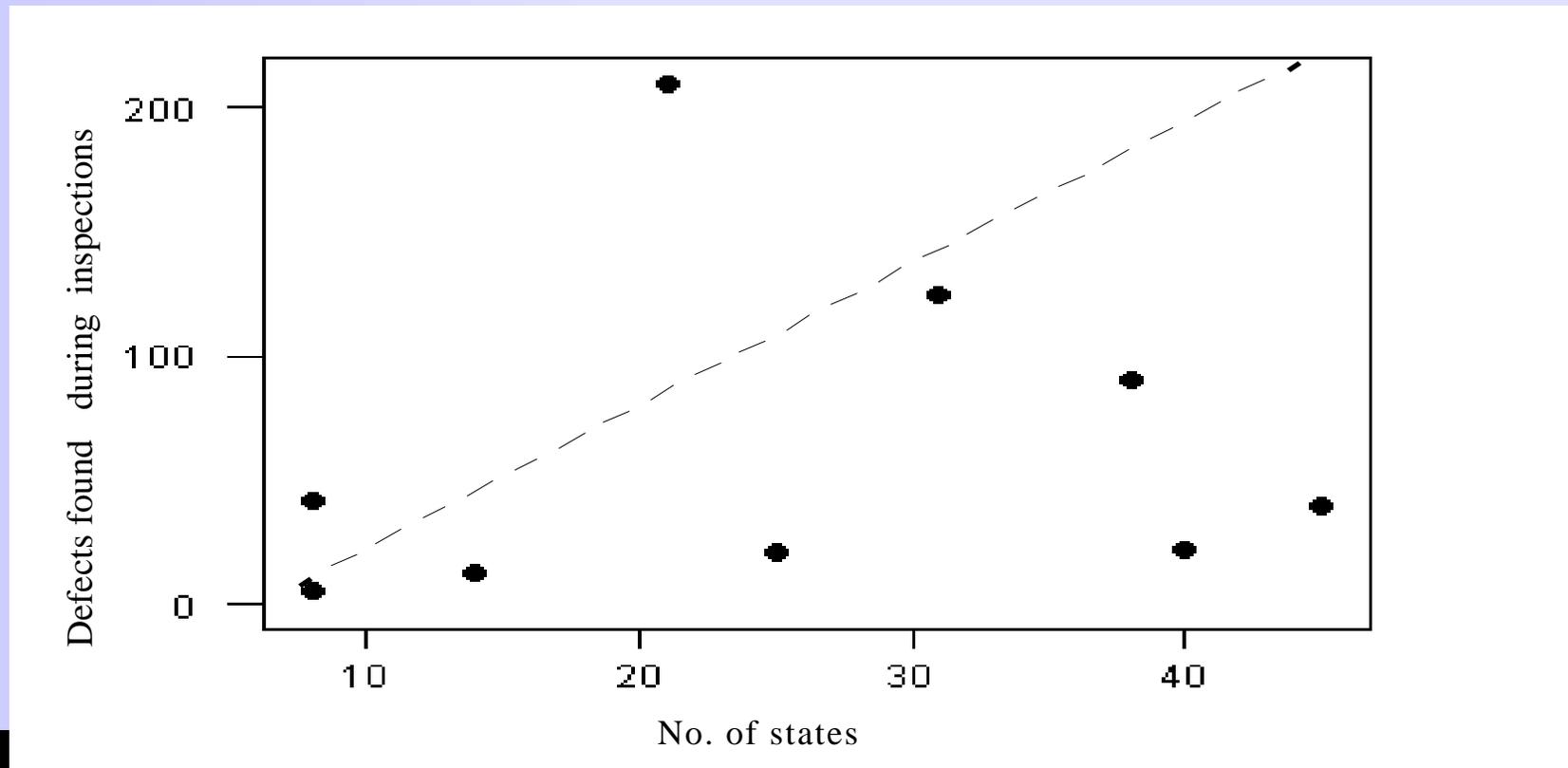
There does not seem to be a well-defined correlation between these two variables. The dashed line shows the expected (intuitively) values.



Results

Study 1: inspection defects and no. of states in a module

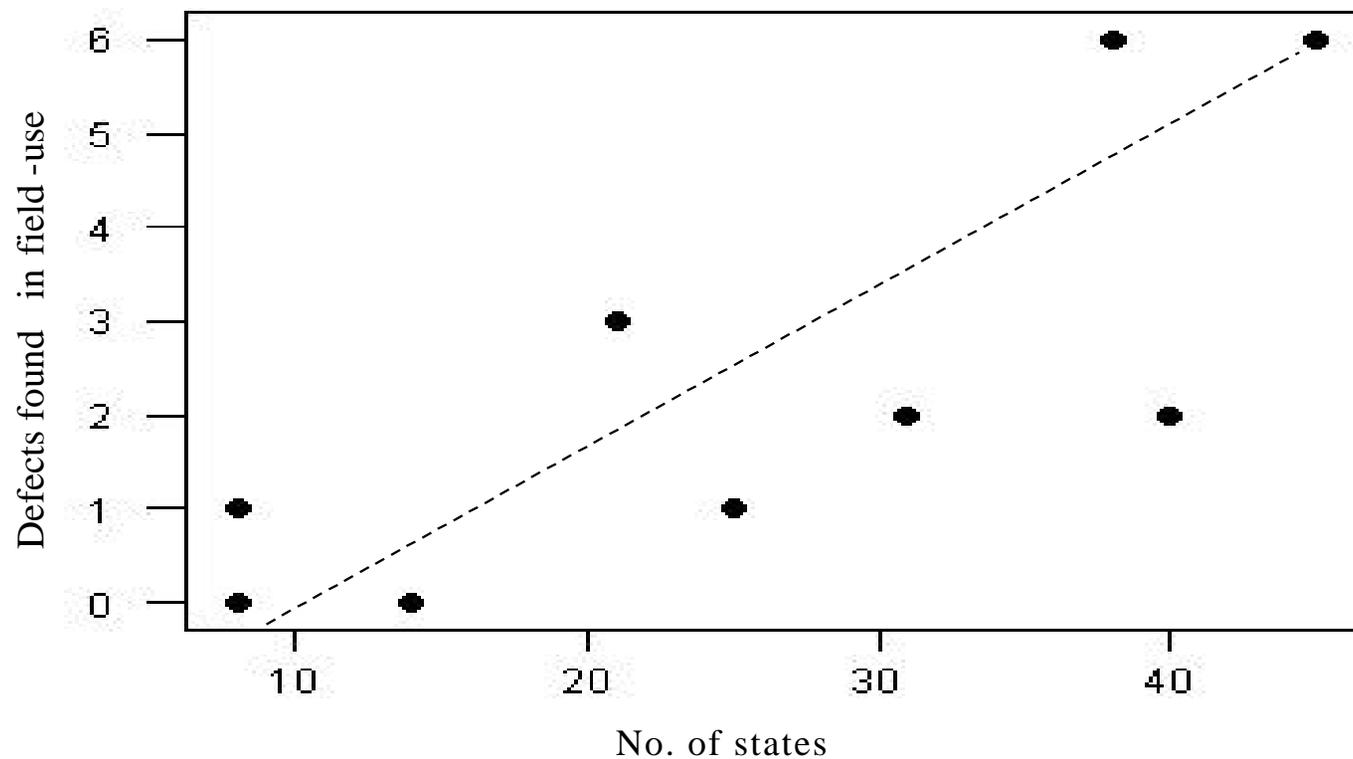
Again, there does not seem to be a well-defined correlation between these two variables. Surprisingly, the no. of defects detected in inspections seems to be rather constant if the topmost value along the y-axis is removed. Intuitively, it should be more difficult to inspect a document with a large number of states, than with a small number of states. The dashed line shows the expected (intuitively) values



Results

Study 1- number of defects in field-use versus states in a module

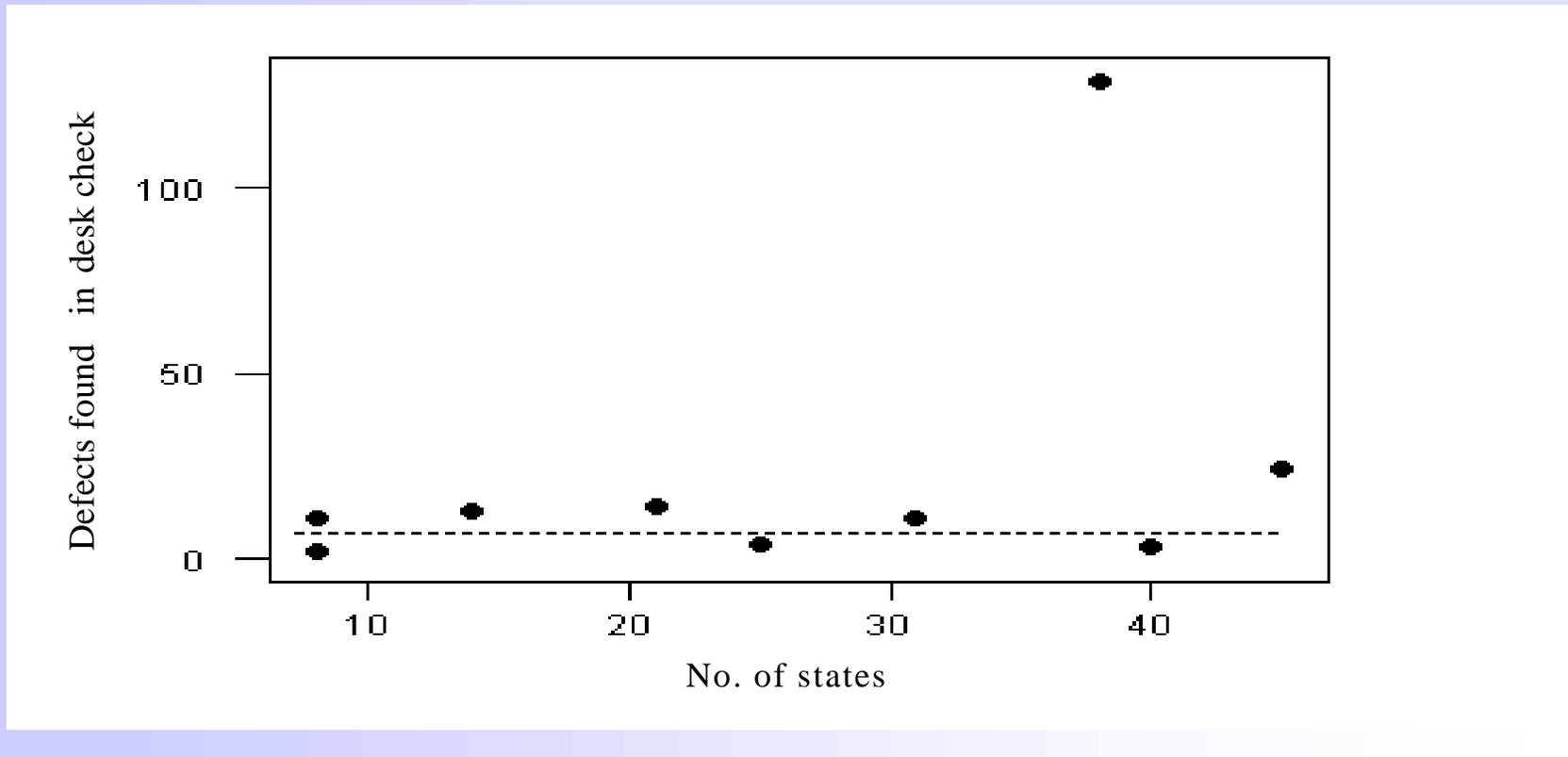
The number of states do seem to be correlated with number of defects in field-use, indicated with the dashed line. The number of system failures increases with increasing number of states. Thus, number of states represent the inherent complexity of a module.



Results

Study 1- defects found in desk check versus states in a module

Surprisingly, the number of defects found in desk check seems to be independent of the number of states in a module. This is indicated in the figure below, if the topmost value along the y-axis is not considered.



Results

Study 1- effort spent on inspections

- Spent: 1474 ph; whereof 1162.5 on individual reading and inspection meetings, and 311.2 of defect fixing.
- Recommended: 2723 ph, according to internal Gilb guidelines.
- Saved: 8200 ph
 - if the defects had not been detected by inspection, but detected and fixed in the later phases (not field-use).
- Inspections detect almost 65% of the registered defects, and desk check 27%. Remaining 7% in later testing activities, and 2% in field-use!!
- At Ericsson, the Gilb inspection process focuses on finding new defects in inspections, but only 3% of the defects found during inspection are actually found in the inspection meeting ("true negatives").
No data stored to verify "false positives".

Results

Study 2- effectiveness of insp. and testing

| Activity | Defects [#] | [%] |
|----------------------------|-------------|-------|
| Inspection reading, design | 4478 | 71.1 |
| Inspection meeting, design | 392 | 6.2 |
| Code review | 832 | 13.2 |
| Unit test | 598 | 9.5 |
| Function test etc. | ? | ? |
| Total | 6300 | 100.0 |

Results

Study 2- cost-effectiveness of insp. & testing

| Activity | Defects [#] | Effort [h] | Effort spent to find one defect [h:m] | Estimated saved effort by early defect removal |
|----------------------------|-------------|------------|---------------------------------------|--|
| Inspection reading, design | 4478 | 5563 | 01:15 | 41000 |
| Inspection meeting design | 392 | 3215 | 08:12 | |
| Code review | 832 | 2440 | 02:56 | - |
| Unit test | 598 | 4388 | 07:20 | - |
| Function test etc. | ? | ? | ? | ? |
| Total | 6300 | 15606 | | |

Results

Study 1 and 2 - Recommended rate vs. actual total effort during inspections, including defect fixing.

Study 1: 1474 ph (54%) actually used,
out of 2723 ph recommended.

Study 2: 20,515 ph (79%) actually used,
out of 26,405 ph recommended.
0.43 defects per page.

Individual reading rate: part of total inspection rate.

All above for design documents, not source code.

Results

Study 2- how cost-efficient are inspection meetings?

Study 2- what kind of defects are found during reading and meetings?

| | Individual reading | Inspection Meeting | Defect fixing | Sum effort |
|------------|--------------------|--------------------|---------------|------------|
| Effort [h] | 5563 | 3215 | 11737 | 20515 |
| [%] | 27.12% | 15.67% | 57.21% | 100.00% |

| | Major defects | | Super Major defects | | Sum defects | Effort | Cost-efficiency |
|-----------------------|---------------|-------|---------------------|------|-------------|--------|-----------------|
| | [#] | [%] | [#] | [%] | [#] | [h] | [h:m/defect] |
| In reading | 4356 | 97.2% | 122 | 2.7% | 4478 | 5563 | 01:15 |
| In meetings | 380 | 96.9% | 12 | 3.1% | 392 | 3215 | 08:12 |
| In defect log (total) | 4736 | 97.2% | 134 | 2.7% | 4870 | 8778 | 01:48 |

Regression Analysis

Study 1- Hypothesis 2

The number of states (N_s) is an important variable, because it correlates to the number of system failures (field-use defects, N_{fu}). The modification rate, N_{mr} , is included in the following regression equation.

$$N_{fu} = \alpha + \beta N_s + \lambda N_{mr}$$

Here α , β , λ are constants.

Regression Analysis - Study 1

- H_2 : the fault density of a module in field-use depends on the complexity of the module (no. of states) and the its modification rate. High complexity and high modification rate will thus result in high fault density in operation.
- H_0 : the fault density of a module in field-use does not depend on the module's complexity and the modification rate.
- H_0 is the null hypothesis, and H_2 is the alternative hypothesis.

H_2 can only be accepted, if β and λ are significantly different from zero and the significance level for each of the coefficients is better than 0.10.

Regression Analysis - Study 1

The following values are estimated:

$$N_{fu} = -1.73 + 0.084N_s + 0.097N_{mr}$$

| Predictor | Coefficient | StDev | t | p |
|-----------------------|-------------|-------|-------|-------|
| Constant (α) | -1.732 | 1.067 | -1.62 | 0.166 |
| States (β) | 0.084 | 0.035 | 2.38 | 0.063 |
| Modrate (λ) | 0.097 | 0.034 | 2.89 | 0.034 |

The estimated values for the coefficients are given above, along with the std. deviations, t-value for testing if the coefficient is 0, and the p-value for this test.
 $S = 1.200$; $R^2 = 79.9\%$, $R^2_{(adj)} = 71.9\%$.

The analysis of variance is summarised below:

| Source | DF | SS | MS | F | P |
|------------|----|-------|-------|------|-------|
| Regression | 2 | 28.68 | 14.34 | 9.96 | 0.018 |
| Error | 5 | 7.20 | 1.44 | | |
| Total | 7 | 35.88 | | | |

It should be noted that the constant is not significant, but the states and modrate are significant. H_2 can be accepted.

Regression Analysis

Study 2- Hypothesis 1

Hypothesis 1 uses the data presented above, and checks whether there exist a correlation between defects found during inspection/test and complexity for a module.

The regression equation used to state this hypothesis can be written as:

$Y = \alpha X + \beta$, where Y is defect density, X is the complexity and α , and β are constants.

H_1 can only be accepted if α and β are significantly different from zero and the significance level for each of the coefficients is better than 0.10.

Regression Analysis

Study 2- Hypothesis 1

- The following values was estimated: $Y = 0.1023 * X + 13.595$

| <i>Predictor</i> | <i>Estimate</i> | <i>Standard error</i> | <i>t</i> | <i>p</i> |
|------------------|-----------------|-----------------------|----------|----------|
| β | 13.595002 | 18.52051 | 0.73 | 0.4729 |
| α | 0.1022985 | 0.093689 | 1.09 | 0.2901 |

It indicates that the linear regression line must be rejected if a significance of level 0.10 is assumed, i.e. *neither H_0 nor H_1 can be refuted*, reason being too few data.

Regression Analysis

Study 2- Hypothesis 3

- To check for correlation between defect densities across phases and deliveries, we have analyzed the correlation between defect densities for modules over two projects.

| | | |
|--------------------------------------|---|---|
| Correlation: 0.4672 | Variable 1: Defect density – Project A | Variable 2: Defect density - Project B |
|--------------------------------------|---|---|

Thus, no significant correlation found between the two data set.
But only 6 modules with complete data for both projects for this test.
So neither H_0 nor H_3 can be refuted.

Conclusions 1(2)

- Inspections are the most cost-effective in defect detection; function testing etc. the least cost-effective.
- Inspections find ca. 70% of recorded defects, takes 6--9% of development effort. **Estimated net saving: 34--21%!!**
- 7% of inspection defects (Study 1: 3% , Study 2: 8%) found during final meeting, 93% during individual reading. But insp. meetings more cost-effective than function test!
- Individual reading & code reviews: the most cost-effective techniques to detect defects.
- Recommended inspection rates not followed, only 54--79% of recommended effort spent.

Conclusions 2 (2)

- Defect density of a module in field-use depends on its complexity and modification rate.
Generally, low-complexity designs have lower defect rates.
But: is it possible to create low complexity designs for real-time telecom software?
Pay extra attention when designing complex parts?
- Ericsson has focussed on finding new defects during meetings; only 7% of inspection-phase defects found here.
- The defect classification is too coarse.
- Must record "false positives"; now only "true negatives".
- Must record overlap between inspectors, to facilitate capture-recapture analysis.

Recommendations

Record and analyze more data properly, e.g. to check if:

- Defect-prone modules during inspections also are defect-prone during later tests and field-use (**longitudinal** analysis)?
- One defect type dominates in one project, or if subsequent projects will have the same defect types?
Will need **better defect classification** (ISO/IEEE)!
- We may **omit inspection meetings** for some document types or try **virtual inspection meetings** on the net/web.