SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# Pitfalls in Epidemiology: From Study to Thesis

Chumpon Wilasrusmee
Bunyarit Sukrat
Boonying Siribumrungwong: Moderator

## PITFALLS

- A hidden or unsuspected danger or difficulty
- A covered pit used as a trap

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

**Epidemiologic studies: pitfalls in interpretation**
Westhoff CL Dialogues Contracept 1995 Winter;4(5):5-6,8

Researchers should design epidemiologic studies in such a way as to avoid or minimize known or suspected biases

Clinicians should consider the aforementioned concerns when interpreting the results of epidemiologic studies

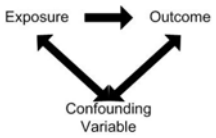
- use well-designed studies and large sample sizes
- Meta-analysis is used increase sample sizes
- In fact, the studies in the meta-analysis tend to be confounders

- They must be prepared to address validity and clinical relevance
- To do so, they need to be familiar with basic study designs and associated issues to provide appropriate counseling and informed clinical decision making

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# OUTLINE

- Pitfalls in development of proposal
- Pitfalls in data collection and management
- Pitfalls in data analysis




Clinical prediction rules

RCT

Real world studies

BIG **data management**

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# PROPOSAL

Pitfall #1: The Presentation

Pitfall #2: The Hazy Question

Pitfall #3: The Inconsistent Protocol

Pitfall #4: Incorrect description of study design

Pitfall #5: The Incorrect Design (or the wrong objective)

Pitfall #6: Randomization/Random sample

Pitfall #7: The Unknown Instrument

Pitfall #8: The Statistical Analysis

Pitfall #9: The Mystery Statistical Consultant

Pitfall #10: Missing Items

- **Role of chance, bias and confounding**
- **Sample size estimation**

http://www.ucalgary.ca/md/CAH/research/pitfalls.htm

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

Example: In this innovative study, I will elucidate, categorize, and illustrate the myriad of characteristics that typify the consumers of emergency health care products. Previous researchers have done a lousy job in view of the fact that their methods were so bad. We also had a notion to study these problems. The findings from our pilot study were great. After a thorough search of the literature, we firmly and honestly believe that an in-depth study which generates a large body of useful information will, in the final analysis, allow better management of emergency room resources in view of the fact that health care resources are stretched thin. **What do you think about this?**

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

- Simplicity
- Clarity
- Parsimony

Avoid jargon.
Avoid trendy words
Avoid abbreviations
Avoid colloquialisms
Do not try to sound "intellectual"
Avoid redundant phrases
Avoid overused phrases

| | |
|---|---|
| • in view of the fact that | **because** |
| • utilize | **use** |
| • elucidate, categorize, and illustrate | **describe** |
| • an in-depth study | **a study** |
| • audible to the ear | **audible** |

Several scoring systems have been developed for diagnosis of appendicitis with interesting results, nevertheless these systems have been less routinely applied in general practice. We had systematically reviewed how those scores were developed and validated, and how their performances were. The review suggested that the research methods for scoring systems of appendicitis were discrepancy. Although there are several diagnostic scoring systems available, applying them to general population might be questionable due to improper methods used for creating scores. The more appropriate scores with internal and external validations are still required

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# QUESTIONS

- The aim of this study is to determine the major concerns of women after a cesarean delivery.
- Patients residing in different parts of Thailand will undergo colonoscopy differently than those residing in other parts.
- Does the administration of analgesic by health care personals vs. by patients themselves affect how older patients feel during postoperative recovery?

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# UNCLEAR

- Make sense?: What kind of concerns? Concepts not identified, Immediately after delivery, a year after delivery?
- Unclear?: what is meant by will undergo and differently? Variables not identified: different parts of Thailand, differently?
- Clear?: Constructs not defined and unmeasurable: feel
- ***Does the administration of narcotic analgesics by health care personales versus patient self-administration affect pain intensity, as measured by the McGill pain score, 24 hours following laparoscopic surgery***.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# RAMA-AS

- What are significant predictors for diagnosis of appendicitis in patients who are suspected of appendicitis?
- What is the performance of RAMA-AS in patients who are suspected of appendicitis?
- Does the  RAMA-AS perfrom better than the previously develop scores?
- Can the RAMA-AS work well in internal validation and external validation?

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

- The purpose of this study is to determine if there are differences in pain control with health care personales versus patient administered analgesia following surgery.
- Research Question: Does the administration of analgesic by nurses vs. by patients themselves affect pain intensity during postoperative recovery in older adults?
- Hypothesis: Patients who self-administered narcotics will be more satisfied than patients who receive narcotics administered by nurses.
- Sample size: To achieve a power of 80% to detect a 20% difference in the total morphine dose in the first 24 hours surgery, 30 subjects in each group will be required.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

## WHAT IS THE PRIMARY OBJECTIVE? DO THE RESEARCHERS KNOW

- **Primary Objectives**

To develop a RAMA-AS for diagnosis of appendicitis in patients who are suspected of appendicitis.

To externally validate RAMA-AS using data from different settings that used for develop score

- **Secondary objectives**

To compare performance of RAMA-AS with the most popular used scoring system, i.e. Alvarado score and previously developed scoring systems

The Inconsistent Protocol?

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

## SAMPLE SIZE

- Common pitfalls The calculation of the sample size is troubled by a large amount of imprecision, because investigators rarely have good estimates of the parameters necessary for the calculation
- Unfortunately, the required sample size is very sensitive to the choice of these parameters

The effects of selecting alpha and the power

Estimating the difference and SD

Post hoc sample calculations

Reporting of sample size calculations

| Your Statistical Decision | True state of null hypothesis | |
|---|---|---|
| | $H_0$ True (example: the drug doesn't work) | $H_0$ False (example: the drug works) |
| Reject $H_0$ (ex: you conclude that the drug works) | *Type I error (α)* | *Correct* |
| Do not reject $H_0$ (ex: you conclude that there is insufficient evidence that the drug works) | *Correct* | *Type II Error (β)* |

| Your Decision | The TRUTH | |
|---|---|---|
| | God Exists | God Doesn't Exist |
| Reject God | *BIG MISTAKE* | *Correct* |
| Accept God | *Correct— Big Pay Off* | *MINOR MISTAKE* |

The effects of selecting alpha and the power

- Type-I Error (also known as "$\alpha$"):
  - Rejecting the null when the effect isn't real.
- Type-II Error (also known as "$\beta$"):
  - Failing to reject the null when the effect is real.
- POWER (the flip side of type-II error: 1- $\beta$):
  - The probability of seeing a true effect if one exists.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# PITFALL: OVER-EMPHASIS ON P-VALUES

- Clinically unimportant effects may be statistically significant if a study is large (result in a small standard error and extreme precision).
- Make attention to effect size and confidence intervals

- A prospective cohort study of 34,079 women found that women who exercised >21 MET hours per week gained significantly less weight than women who exercised <7.5 MET hours (p<.001)
- Conclusion: "To Stay Trim, Women Need an Hour of Exercise Daily."

**Table 2.** Mean (SD) Differences in Weight Over Any 3-Year Period by Physical Activity Level, Women's Health Study, 1992

| Group | No. of Women[b] | Physical Activity, MET Hours per Week | | | *P* Value for Trend |
|---|---|---|---|---|---|
| | | <7.5 | 7.5 to <21 | ≥21 | |
| All women | | | | | |
| Analytical model[c] | | | | | |
| 1 | | 0.15 (0.04) | 0.12 (0.04) | 0 [Reference] | <.001 |

- What was the effect size? Those who exercised the least 0.15 kg (.33 pounds) more than those who exercised the most over 3 years.
- Analysis: extrapolated over 13 years of the study, the high exercisers gained 1.4 pounds less than the low exercisers!
- *Great example of a statistically significant effect that is not clinically significant.*

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

## PITFALL: ASSOCIATION DOES NOT EQUAL CAUSATION

- Key point: Statistical significance does not imply a cause-effect relationship.
- Recommendation: Interpret results in the context of the study design.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# PITFALL: MULTIPLE COMPARISONS

- A significance level of 0.05 implys that your false positive rate for one test is 5%.
- If you test more than one test, your false positive rate will be higher than 5%.

-Look at the totality of the evidence.
-Expect about one marginally significant p-value (.01<p or p<.05) for every 20 tests run.
-Be aware of unplanned comparisons (subgroup analyses).

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

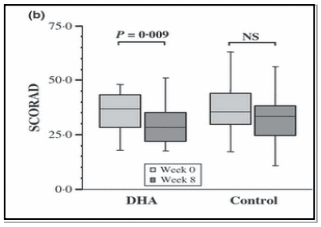
# PITFALL: HIGH TYPE II ERROR (LOW STATISTICAL POWER)

- Results that are not statistically significant should not be interpreted as "evidence of no effect," but interprete as "no evidence of effect"
- Studies may miss effects if they are insufficiently powered (lack precision).
- Example: A study of 63 postmenopausal women failed to find a significant relationship between hormone replacement therapy and prevention of vertebral fracture. The odds ratio and 95% CI were: 0.38 (0.12, 1.19), indicating a potentially meaningful clinical effect.
- Failure to find an effect may be caused by insufficient statistical power for this endpoint.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# PITFALL: THE FALLACY OF COMPARING STATISTICAL SIGNIFICANCE

- "The effect was significant in the treatment group, but not significant in the control group" does not mean that the groups differ significantly
- Statistical power is the probability of finding an effect if it's true (real)




Also applies to interactions…

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# FACTORS AFFECTING POWER

1. Size of the effect: Increase
2. Standard deviation of the characteristic: Decrease
3. Bigger sample size: Increase
4. Significance level desired: Decrease

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

## SIMPLE FORMULA FOR DIFFERENCE IN PROPORTIONS

Sample size in each group (assumes equal sized groups)

Represents the desired power (typically .84 for 80% power).

**Sample size** $$n = 2 \times \frac{(\bar{p})(1-\bar{p})(Z_{\beta} + Z_{\alpha/2})^2}{(p_1 - p_2)^2}$$

A measure of variability (similar to standard deviation)

Effect Size (the difference in proportions)

Represents the desired level of statistical significance (typically 1.96).

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

## SIMPLE FORMULA FOR DIFFERENCE IN MEANS

Sample size in each group (assumes equal sized groups)

Represents the desired power (typically .84 for 80% power).

**Sample size** $$n = 2 \times \frac{\sigma^2 (Z_{\beta} + Z_{\alpha/2})^2}{\text{difference}^2}$$

Standard deviation of the outcome variable

Effect Size (the difference in means)

Represents the desired level of statistical significance (typically 1.96).

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# RAMA-AS

- As for our pilot study, proportion appendicitis in exposure and non-exposure groups for variables that associated with appendicitis are described in Table 3.1.
- Type I and type II errors are set at 5% and 20%, with the ratio of exposure vs non-exposure of 1: 1.
- The size of detectable is set at 5% and 10%.
- Sample size estimation from each variable was calculated. Using data from the most significant variable (migration of pain).
- A total of 8-10 parameters were expected to be included in the final score. Using the rule of thumb, 20 subjects with appendicitis were required for 1 variable, approximately 160 and 200 subjects with appendicitis were needed for 8 and 10 parameters.
- Comparing estimated sample sizes between the two approaches, a larger sample size was used. Given percentage of loss to follow up of 5%.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# PITFALL

- Numerous systematic reviews have observed that prediction model studies, both development and validation studies, frequently provided no rationale for the sample size or any mention of overfitting

"We did not calculate formal sample size calculations because all the cohort studies are ongoing studies. Also there are no generally accepted approaches to estimate the sample size requirements for derivation and validation studies of risk prediction models. Some have suggested having at least 10 events per candidate variable for the derivation of a model and at least 100 events for validation studies"

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

## TRANSPARENT REPORTING OF A MULTIVARIABLE PREDICTION MODEL FOR INDIVIDUAL PROGNOSIS OR DIAGNOSIS (TRIPOD)

- Although there is a consensus on the importance of having an adequate sample size for developing a prediction model, how to determine what counts as “adequate” is not clear.
- As for all medical research, a larger sample size yields more precise results.
- In the absence of bias, larger samples also yield more reliable findings.
- Crucially, in prediction studies (development and validation), the number of outcome events dictates the effective sample size

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

## DEVELOPMENT STUDY

- Model's performance is likely to be overestimated when it is developed and assessed for its predictive accuracy on the same data set
- A rule of thumb for sample size was suggested that has been quite widely adopted.
- The rule is to have at least 10 outcome events per variable (EPV), or more precisely, per parameter estimated.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# VALIDATION STUDY

- Sample size requirements for validation studies are not well understood, and there is a dearth of empirical evidence to guide investigators.
- The limited empirical evidence to support investigators in guiding their sample size choice for validation studies suggests a minimum of 100 events and 100 nonevents, whereas more than 250 events have been suggested as preferable.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# DATA COLLECTION AND MANAGEMENT

- Equipment failure, environmental hazards, and transcription errors
- Lack of internal consistency – purpose isn't met by design, instruments or methods won't result in answers to the question OR – a+ b can't add up to c
- Not enough data or those that you have are not convincing/credible/ well organized: same as: No "golden thread" each section has its own focus but does not tie back to the focus of your study
- Do's Develop and describe the findings of the thesis thoroughly so that you are completely credible to the reader

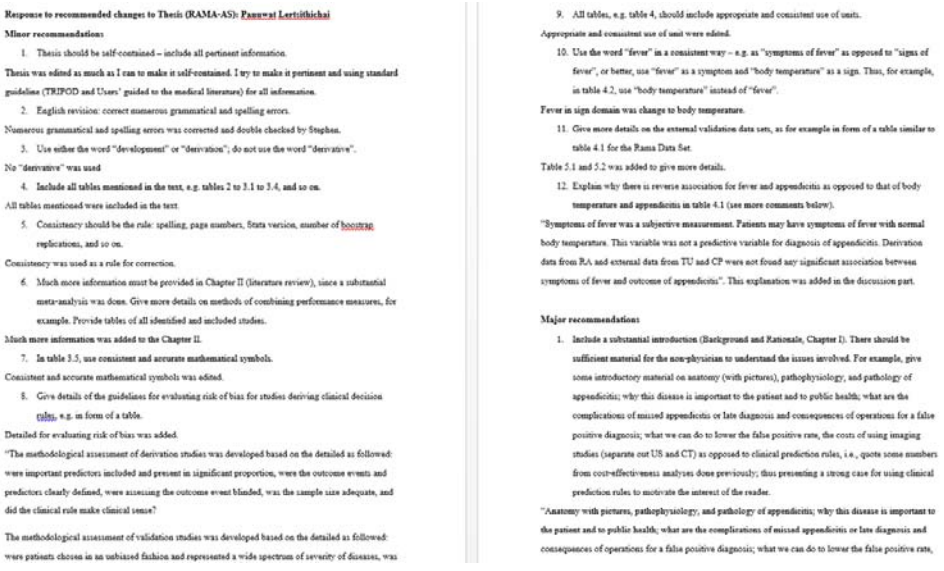
SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# DATA ANALYSIS

- Use statistics without confident that your analysis answers the questions you are asking
- Treat a write-up like a diary, with EVERYTHING in it rather than just what worked or was properly designed
- "Misses the plot" Qualitative • Doesn't answer the question • No clear path from data to results • Leaves us asking questions Quantitative • Misuse of statistical measures (descriptive, inferential, comparative, relational what are you using and why?) • Your reader should not have to understand statistics to understand your findings or results • Measures used for no apparent reason

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# RAMA-AS: EXTERNAL VALIDATION



**Response to recommended changes to Thesis (RAMA-AS): Panuwat Lertsithichai**

**Minor recommendations:**

1. Thesis should be self-contained – include all pertinent information.

Thesis was edited as much as I can to make it self-contained. I try to make it pertinent and using standard guideline (TRIPOD and Users' guided to the medical literature) for all information.

2. English revision: correct numerous grammatical and spelling errors.

Numerous grammatical and spelling errors was corrected and double checked by Stephen.

3. Use either the word "development" or "derivation"; do not use the word "derivative".

No "derivative" was used

4. Include all tables mentioned in the text, e.g. tables 2 to 3.1 to 3.4, and so on.

All tables mentioned were included in the text.

5. Consistency should be the rule: spelling, page numbers, Stata version, number of bootstrap replications, and so on.

Consistency was used as a rule for correction.

6. Much more information must be provided in Chapter II (literature review), since a substantial meta-analysis was done. Give more details on methods of combining performance measures, for example. Provide tables of all identified and included studies.

Much more information was added to the Chapter II.

7. In table 3.5, use consistent and accurate mathematical symbols.

Consistent and accurate mathematical symbols was edited.

8. Give details of the guidelines for evaluating risk of bias for studies deriving clinical decision rules, e.g. in form of a table.

Detailed for evaluating risk of bias was added.

"The methodological assessment of derivation studies was developed based on the detailed as followed: were important predictors included and present in significant proportion, were the outcome events and predictors clearly defined, were assessing the outcome event blinded, was the sample size adequate, and did the clinical rule make clinical sense?

The methodological assessment of validation studies was developed based on the detailed as followed: were patients chosen in an unbiased fashion and represented a wide spectrum of severity of diseases, was

9. All tables, e.g. table 4, should include appropriate and consistent use of units.

Appropriate and consistent use of unit were edited.

10. Use the word "fever" in a consistent way – e.g. as "symptoms of fever" as opposed to "signs of fever", or better, use "fever" as a symptom and "body temperature" as a sign. Thus, for example, in table 4.2, use "body temperature" instead of "fever".

Fever in sign domain was change to body temperature.

11. Give more details on the external validation data sets, as for example in form of a table similar to table 4.1 for the Rama Data Set.

Table 5.1 and 5.2 was added to give more details.

12. Explain why there is reverse association for fever and appendicitis as opposed to that of body temperature and appendicitis in table 4.1 (see more comments below).

"Symptoms of fever was a subjective measurement. Patients may have symptoms of fever with normal body temperature. This variable was not a predictive variable for diagnosis of appendicitis. Derivation data from RA and external data from TU and CP were not found any significant association between symptoms of fever and outcome of appendicitis". This explanation was added in the discussion part.

**Major recommendations:**

1. Include a substantial introduction (Background and Rationale, Chapter I): There should be sufficient material for the non-physician to understand the issues involved. For example, give some introductory material on anatomy (with pictures), pathophysiology, and pathology of appendicitis; why this disease is important to the patient and to public health; what are the complications of missed appendicitis or late diagnosis and consequences of operations for a false positive diagnosis; what we can do to lower the false positive rate, the costs of using imaging studies (separate out US and CT) as opposed to clinical prediction rules, i.e., quote some numbers from cost-effectiveness analyses done previously; thus presenting a strong case for using clinical prediction rules to motivate the interest of the reader.

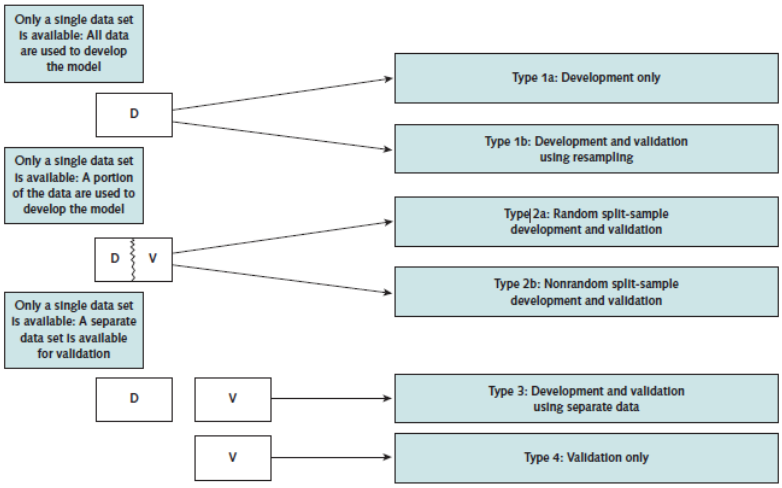
"Anatomy with pictures, pathophysiology, and pathology of appendicitis; why this disease is important to the patient and to public health; what are the complications of missed appendicitis or late diagnosis and consequences of operations for a false positive diagnosis; what we can do to lower the false positive rate,

There should be a substantial discussion on the differences between the 3 data sets from the 3 institutions

The more serious, and I think the most troubling part of the thesis is the large weight given to the body temperature variable.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# RAMA-AS




SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# TRIPOD

- For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors, and outcome)
- Authors should explain the reasons for any notable differences between the validation samples and the previous study samples, and later in the article consider the possible implications on the found results, such as the model's predictive performance in the validation set.
- In a recent systematic review of 78 external validation studies (including development studies with an external validation), only 31 (40%) compared or discussed the characteristics of both the original development and external validation cohorts

*Annals of Internal Medicine*: 162 (1): W1-W73

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# CHARACTERISTICS

| Characteristics | RA (n=396) | TS (n=152) | CP (n=178) |
|---|---|---|---|
| Mean age (SD),years | 36.3(14.6) | 35.6(16.9) | 42.9(16.8) |
| Men | 132 (35.8%) | 40 (26.4%) | 71 (39.9%) |
| Symptoms | | | |
| Progression of pain | 336 (84.8%) | 98 (64.5%) | 147 (82.6%) |
| Aggravation of pain | 287 (72.5%) | 78 (51.4%) | 104 (58.4%) |
| Migration of pain | 177 (44.7%) | 73 (48.0%) | 125 (70.2%) |
| Signs | | | |
| Body temperature ≥37.8 °C | 74 (18.7%) | 30 (19.7%) | 67 (37.6%) |
| Rebound tenderness | 192 (48.5%) | 65 (42.8%) | 127 (71.3%) |
| Laboratory | | | |
| WBC | 315 (79.6%) | 125 (82.2%) | 141 (79.2%) |
| Neutrophil | 262 (66.2%) | 115 (75.7%) | 124 (69.7%) |
| Prevalence of appendicitis | 245/396 (61.8%) | 74/152 (48.7%) | 137/178 (76.9%) |

-More rebound tenderness is elicited in CP

-

Prevalence higher in

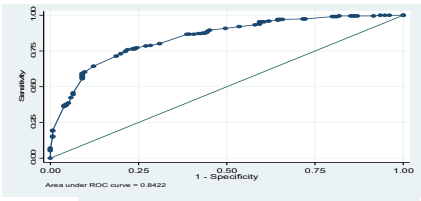
# RAMA-AS

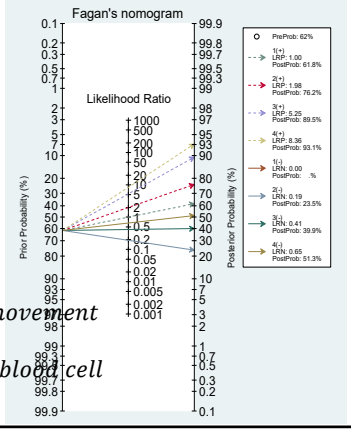
SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

| Score (Variables) | Risk groups | Score development for derivative phase | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Outcome | | %Sensitivity (95%CI) | %Specificity (95%CI) | LR+ (95%CI) | LR- (95%CI) | Post-positive test odds (%) |
| | | AP | Non-AP | | | | | |
| <-0.64 | Very low risk | 25 | 85 | 100.00 | 0 | 1.00 | 0 | 61.80 |
| -0.64 to 0.84 | Low risk | 61 | 51 | 89.75 (85.25-93.26) | 54.97 (46.67-63.06) | 1.98 (1.65-2.37) | 0.19 (0.13-0.28) | 76.00 (73.00-79.00) |
| 0.85 to 1.74 | Moderate risk | 64 | 12 | 64.08 (57.73-70.09) | 88.08 (81.82-92.78) | 5.25 (3.39-8.13) | 0.41 (0.34-0.49) | 89.00 (85.00-93.00) |
| >1.74 | High risk | 91 | 7 | 37.96 (31.86-44.36) | 95.36 (90.68-98.12) | 8.36 (3.96-18.00) | 0.65 (0.59-0.72) | 93.00 (86.00-97.00) |

| Symptoms | Progression of pain |
|---|---|
| | Aggravation of pain by cough or movement |
| | Migration of pain |
| Signs | Body temperature ≥37.8 °C |
| | Rebound tenderness |
| Lab results | WBC >10,000 cell/mm$^3$ |
| | Neutrophil >75% |

$$ln[P/(1-P)] = -3.37 + (0.80)migration\ of\ pain + (1.04)progression\ of\ pain + (0.78)aggravation\ of\ pain\ by\ cough\ or\ movement + (1.64)Body\ temperature + (1.53)rebound\ tenderness + (0.91)white\ blood\ cell + (0.69)neutrophil$$

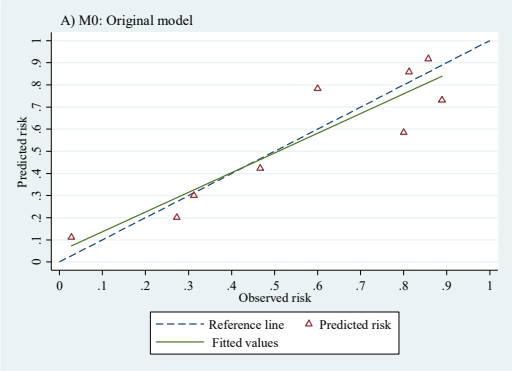


SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

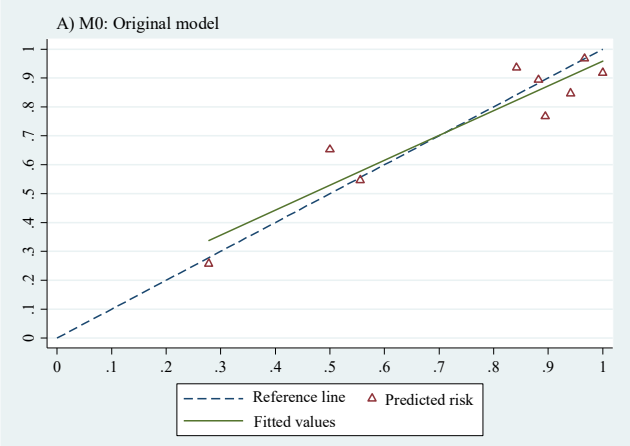
## APPLYING AND QUANTIFYING THE MODEL'S PREDICTIVE PERFORMANCE

- **TU**, the estimated RAMA-AS ranged -3.4 to 4.1 with a median of 0.2
- The derivative model seemed to work well, with the estimated O/E ratio of 1.01 (95%CI:0.78, 1.23)




**However, the calibration plot showed the predicted risk deviated from the reference line i.e. over-estimated risk for lower score and under estimated risk for higher score.**

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University




- **Chaiyaphum Hospital**, a median RAMA-AS was 1.7 (-3.4, 4.1) with O/E ratio of 0.99 (95% CI: 0.65, 1.33
- M0 still deviated from the reference line particularly for low and high scores

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# MODEL UPDATING

- If done, report the results from any model updating (i.e., model specification, model performance)
- The performance of an existing model with new data is often poorer than in the original development sample.
- The investigators may then decide to update or recalibrate the existing model in one of several ways
- Depending on the method of updating, this includes reporting the reestimated intercept, updated regression coefficients (for example, using the slope of the calibration plot of the original model in the validation set), or the estimated regression coefficients of the model, including any new predictors.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# RATIONALE

*Heart.BMJ.com August 2012*

| Method | Updating method | Reason for updating |
|---|---|---|
| 0 | No adjustment (the original prediction model) | – |
| 1 | Adjustment of the intercept (baseline risk) | Difference in the outcome frequency (prevalence or incidence) between development and validation sample |
| 2 | Method 1 + adjustment of all predictor regression coefficients by one overall adjustment factor | Regression coefficients of the original model are overfitted (or underfitted) |
| 3 | Method 2 + extra adjustment of regression coefficients for predictors with different strength in the validation sample as compared with the development sample | As in method 2, and the strength (regression coefficient) of one or more predictors may be different in the validation sample |
| 4 | Method 2 + stepwise selection of additional predictors | As in method 2, and one or more potential predictors were not included in the original model, or a newly discovered marker may need to be added |
| 5 | Re-estimation of all regression coefficients, using the data of the validation sample only | The strength of all predictors may be different in the validation sample, or the validation sample is much larger than the development sample |
| 6 | Method 5 + stepwise selection of additional predictors | As in method 5, and one or more potential predictors were not included in the original model |

**Model updating**

- Objective: To adjust and/or improve the performance of an existing model for other institutions, countries, clinical settings or individual patient populations.
- Indication: Poor performance of the original model in an external validation study.
- Requirements: Ideally, individual participant data from the new situation.
- Methods: Updating methods range from simple adjustment of the baseline risk/hazard, to additional adjustment of predictors weights using the same or different adjustment factors, to re-estimating predictor weights and adding new predictors or removing existing predictors from the original model.
- Model performance: Successfully updating an existing model can result in improved calibration alone, or in improved calibration and discrimination in the new situation, depending on the extent of model adjustment/updating.
- Further validation: Just like a newly developed prediction model, adjusted or updated models should ideally also go through external validations.

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

- **Re-calibrations were performed by re-calibrate intercept (called M1) and overall coefficient (called M2)**

| Type of update model | Thammaasat | Chaiyaphum |
|---|---|---|
| $M_0$: Original model | | |
| $\hat{\alpha}_0$ | $\hat{\alpha}_1 + \hat{\beta}_1 x(RAMA-AS)$ | $\hat{\alpha}_2 + \hat{\beta}_2 x(RAMA-AS)$ |
| M1: Re-calibrate intercept | | |
| α | $\hat{\alpha}_0 \pm \hat{\alpha}_1$ | $\hat{\alpha}_0 \pm \hat{\alpha}_2$ |
| M2: Re-calibration | | |
| α | $\hat{\alpha}_0 \pm \hat{\alpha}_1$ | $\hat{\alpha}_0 \pm \hat{\alpha}_2$ |
| $\beta_{overall}$ | $\hat{\alpha}_0 \pm \hat{\alpha}_1 + \hat{\beta}_1(\Sigma \hat{\beta}_i x_i)$ | $\hat{\alpha}_0 \pm \hat{\alpha}_2 + \hat{\beta}_2(\Sigma \hat{\beta}_i x_i)$ |

- **The M1 was constructed by fitting RAMA-AS on appendicitis variable in two external data separately**
- **The estimated intercept of this model was then used to re-calibrate by minus or plus depending on prevalence of appendicitis from the original intercept**

- **M2: An estimated coefficient from this M1 model was then used to calibrate coefficient by multiplying it with overall coefficients**

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

| | | |
|---|---|---|
| M3: Revision $M2+\gamma_i X_i$ | | |
| α | $\hat{\alpha}_0 \pm \hat{\alpha}_1$ | $\hat{\alpha}_0 \pm \hat{\alpha}_2$ |
| $\beta_{overall}$ | $\hat{\alpha}_0 \pm \hat{\alpha}_1 + \hat{\beta}_1(\Sigma \hat{\beta}_i x_i)_i$ | $\hat{\alpha}_0 \pm \hat{\alpha}_2 + \hat{\beta}_2(\Sigma \hat{\beta}_i x_i)$ |
| Likelihood ratio test | $+\hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \ldots$ | $+\hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \ldots$ |
| M4: Revision $M2+\gamma_i X_i$ | | |
| A | $\hat{\alpha}_0 \pm \hat{\alpha}_1$ | $\hat{\alpha}_0 \pm \hat{\alpha}_2$ |
| $\beta_{overall}$ | $\hat{\alpha}_0 \pm \hat{\alpha}_1 + \hat{\beta}_1(\Sigma \hat{\beta}_i x_i)$ | $\hat{\alpha}_0 \pm \hat{\alpha}_2 + \hat{\beta}_2(\Sigma \hat{\beta}_i x_i)$ |
| Stepwise selection | $+\hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \ldots$ | $+\hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \ldots$ |
| M5: | | |
| Enter all predictors | $\hat{\alpha}_1 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_p x_p$ | $\hat{\alpha}_2 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_p x$ |
| M6: | | |
| Stepwise selection | $\hat{\alpha}_1 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots$ | $\hat{\alpha}_2 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots$ |

- In addition, four model revisions were additionally performed from the M2 as follows
- The M3 was constructed by fitting M2 with additional individual of seven predictors.
- A likelihood ratio test was applied to compare the M2 with/without additional predictor.
- Only significant predictors were kept in the model M3.
- The M4 was constructed by fitting M2 plus significant predictors from stepwise selections of seven predictors.
- The M5 was constructed by re-estimate all coefficients of seven predictors.
- Finally, the M6 was constructed by re-select only significant predictors out of seven predictors

$\hat{\alpha}_0 \pm \hat{\alpha}_1$

| Type of update model | Thammasat | Chaiyaphum |
|---|---|---|
| $M_0$: Original model | | |
| α | -3.397 | -3.397 |
| M1: Re-calibration | | |
| α | -3.397-0.231 | -3.397-(-0.578) |
| M2: Re-calibration | | |
| α | -3.397-0.231 | -3.397-(-0.578) |
| $\beta_{overall}$ | 0.915 | 0.779 |
| M3: Revision M2+$\gamma_i X_i$ | | |
| α | -3.397-0.231 | -3.397-(-0.578) |
| $\beta_{overall}$ | 0.915 | 0.779 |
| Migration of pain | 1.332 | |
| Progression of pain | 1.219 | |
| Fever | -1.247 | -1.259 |
| Rebound tenderness | | 1.976 |
| WBC | -1.476 | |
| Neutrophil | 1.383 | -1.413 |
| M4: | | |
| A | -3.397-0.231 | -3.397-(-0.578) |
| $\beta_{overall}$ | 0.915 | 0.779 |
| Migration of pain | 1.836 | 0.957 |
| Progression of pain | 1.768 | 1.128 |
| Rebound tenderness | 1.817 | 2.619 |
| M5: | | |
| α | -3.189 | -1.965 |
| Migration of pain | 1.816 | 1.011 |
| Aggravation of pain | 0.512 | 0.988 |
| Progression of pain | 1.539 | 0.197 |
| Fever | 0.485 | 0.405 |
| Rebound tenderness | 1.720 | 2.566 |
| WBC | 0.345 | 1.1286 |
| Neutrophil | -0.282 | -0.757 |
| M6: | | |
| α | -2.959 | -1.430 |
| Migration of pain | 1.836 | 0.957 |
| Progression of pain | 1.768 | |
| Aggravation of pain | | 1.128 |
| Rebound tenderness | 1.817 | 2.619 |

M1: fitting RAMA-AS on variable in external datasets separately. The estimated intercept used to re-calibrate

M2: estimated coefficient from this M1 model was then used to calibrate coefficient by multiplying it with overall coefficients

M3: constructed by fitting M2 with additional individual of significant predictors suggested from the derive model

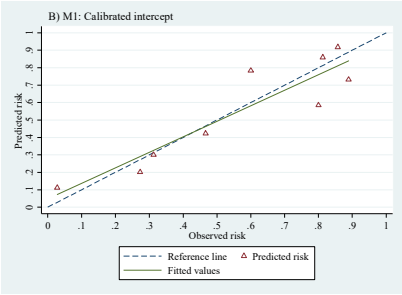
M4: constructed by fitting M2 plus significant predictors from stepwise selections of predictors included in the original model

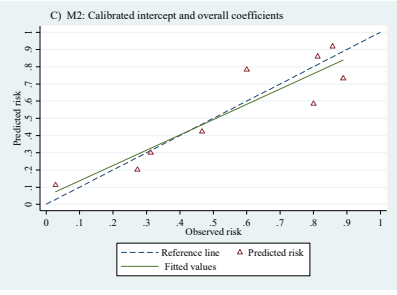
M5: constructed by re-estimate all coefficients of predictors in the original model

M6: re-select only significant predictors out of all the predictors included in the original model

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

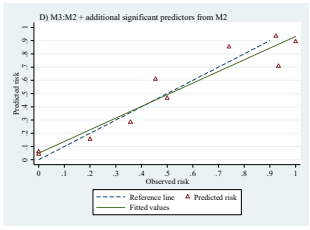
- Calibration plots were constructed which suggested no improvement of calibrations

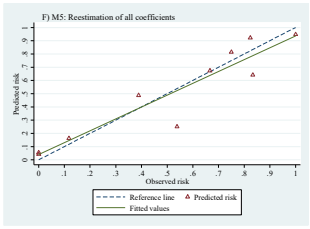


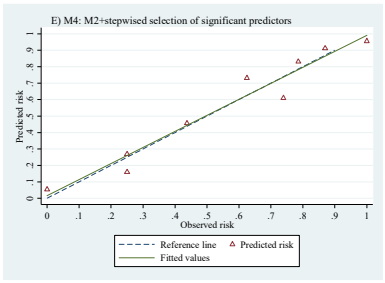



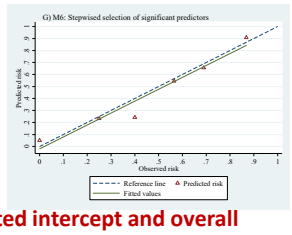
**SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS**
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

- Revision methods of M4-M6 were then constructed, calibrations of these models were plotted



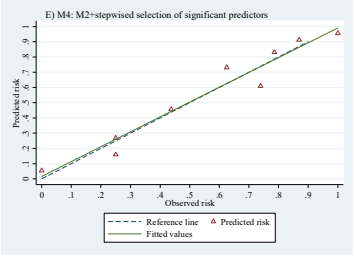




**These suggested that the O/E ratio for revision model M4 (calibrated intercept and overall coefficients plus stepwise selection of significant predictors) was 0.960 (95% CI: 0.750, 1.160; Hosmer-Lemshow = 6.00, df = 7, p = 0.539), which was much improved when compared to M0**

**SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS**
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

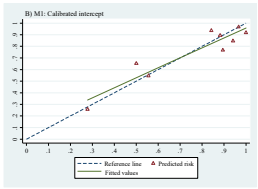
- C-statistics were estimated for all models
- This suggested that the RAMA-AS could well discriminate appendicitis from non-appendicitis with the C-statistics of 0.840 (95%CI: 0.780, 0.910), and 0.879 (95% CI: 0.825, 0.933) for M0 and M4

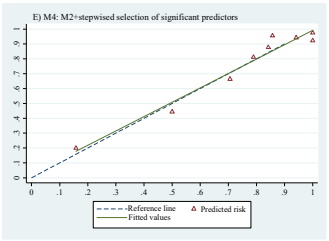
| Data set | Thammasat Hospital |
|---|---|
| M0 | 0.844 (0.780, 0.908) |
| M1 | 0.844 (0.780, 0.908) |
| M2 | 0.844 (0.780, 0.908) |
| M3 | 0.877 (0.823, 0.931) |
| **M4** | **0.879 (0.825, 0.933)** |
| M5 | 0.884 (0.832, 0.935) |
| M6 | 0.872 (0.817, 0.926) |

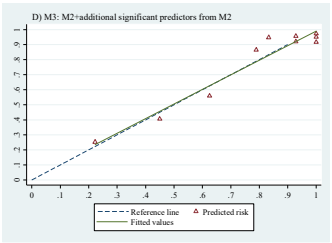


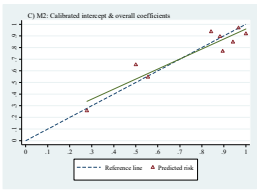
- Calibration of intercept and overall coefficient still did not improve calibration when compared to original M0.

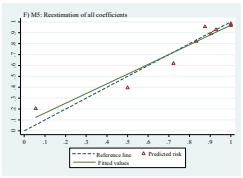
**SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS**
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

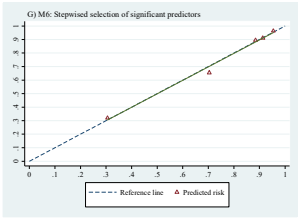



**Among revision of M3-M6 models, M3, M4, and M6 were very much improved in calibrations with O/E ratios of 1.000 (95% CI: 0.940, 1.070), 0.990 (95% CI: 0.910, 1.070), 1.000 (95% CI: 0.950, 1.050)**

- The estimated C-statistics for M0, M3, M4, and M6 were 0.810 (95%CI: 0.730, 0.890), 0.858 (95% CI: 0.788, 0.928), 0.857 (95% CI: 0.788, 0.925), 0.860 (95% CI: 0.790, 0.930)

**SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS**
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

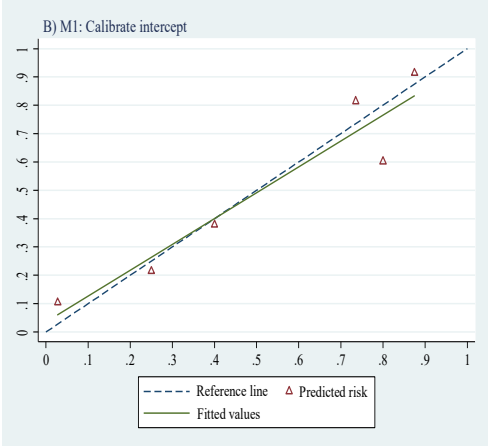
| Data set | Chaiyaphum Hospital |
|---|---|
| M0 | 0.809 (0.728, 0.889) |
| M1 | 0.809 (0.728, 0.889) |
| M2 | 0.809 (0.728, 0.889) |
| M3 | 0.858 (0.788, 0.928) |
| M4 | 0.859 (0.789, 0.928) |
| M5 | 0.873 (0.809, 0.938) |
| M6 | 0.860 (0.790, 0.930) |

- **External validation**
- gen score = -3.374991 + .7978508*symptom3 + 1.042774*symptom5 + .7787298*movecoug + 1.529419*rebten + 1.636311 *fever2 + .9095179*labwbc2 + .6898424*labneu2
- sum score
- logit appendicitis score
- lroc
- roctab appendicitis score
- predict p, pr
- estat gof, gr(6)
- A) Recalibrate intercept
- Recalibrate the constant term due to difference of incidence/prevalence of appendicitis between derive and validate data
- Correction factor = ln[{(prev/(1-prev)}/{(MPV/(1-MPV)]
- MPV=mean predicted value, in which would be exactly the same as the prevalence of disease given predicted values = estimated P from predict command.
- Thus, I use median instead
- sum p,det
- 50% .4218531 Mean .4868421
- disp ln((.4868421/(1-.4868421))/(.4218531/(1-.4218531)))
- .26252708
- gen cf = ln((.4868421/(1-.4868421))/(.4218531/(1-.4218531)))

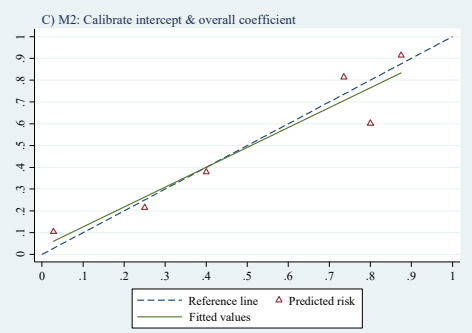
***Prevalence of appendicitis in the derive data
M1 Calibration of intercept (baseline risk)
If the prev(original) is higher than prev(val), decrease prev(org) is required by add up CF because it is +cf
calibrate b0 = b0(orig) - cf ; if cf > 0, b0-cf
If the prev(orig) is lower than prev(val), increasing prev(org) can be done by subtraction the cf from the b0
calibrate b0 = b0(orig) + cf
For appendicitis data, the prev(org) < prev(val),
thus,
calb0 = b0(org) - cf

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

***M1
logit app score
gen score1 = (-3.374991-cf) + .7978508*symptom3 + 1.042774*symptom5 + .7787298*movecoug + 1.529419*rebten + 1.636311 *fever2 +.9095179*labwbc2 + .6898424*labneu2
***Refit the equation using cal_b0 as below. this model also yields calibrated coefficient (cal_b), which can be used for the next step of calibration coefficient
ln [risk of app/(1 - risk of app)] = cal_b0 + cal_b*xb
logit append score1
lroc
roctab appen score1
***Calibration for m1
logit append score1
estat gof, gr(6) tab
disp chiprob(4,8.219118) /*df = 6-1-1*/
***Calibration plot
twoway (line obs_p1 ref p1 ref, lpattern(dash)) (scatter p1 obs_p1 , sort msymbol(triangle_hollow) ylabel(0(.1)1) xlab(0(.1)1)) (lfit p1 obs_p1)
ci mean o e

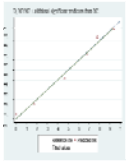


- M2: Calibration by correction of all coefficients by one overall correction factor
- i.e., cal_b as defined below. This is because original coefficients are overfitted or underfitted.
- As for ln [risk of app/(1 - risk of app)] = cal_b0 + cal_b*xb
- logit app score
- gen cal_b = _b[score]
- sum cal_b
- use the cal_b to calibrate coefficients in the original mo
- gen score2 = (-3.374991-cf) + cal_b*(.7978508*symptom3 + 1.042774*symptom5 + .7787298*movecoug + 1.529419*rebten + 1.636311 *fever2 + .9095179*labwbc2 + .6898424*labneu2)
- logit app score2
- roctab appen score2
- ***Calibration for M2
- ci mean o_e

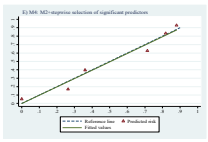





- M3: Update model or revision method
- **This is M2 + additional adjustment of regression coefficient for few predictors which are under/overfitted in validate data compared to original data**
- **ln [risk of app/(1 - risk of app)] = cal_b0 + cal_b*xb + gamma*predictor**
- **refit the equation the same as calibration coefficient but include each predictor one by one in this model.**
- **If gamma coefficient is significant, that means that predictor is needed to calibrate**
- Compare [cal_b0+cal_b*xb + gammaX] vs [cal_b0+cal_b*xb]
- logit app score2 /*cal_b0 + cal_b*xb*/
- estimates store cal_b0b
- logit app score2 /*cal_b0 + cal_b*xb*/
- ***keep LL for M2
- estimates store cal_b0b
- foreach var of varlist symptom3 symptom5 move fever2 rebten labwbc2 labneu2 {
- logit append score2 `var'
- estimates store m`var'
- lrtest cal_b0b m`var'
- }



- As a result, this comparisons between the two models, with vs without that predictor
- indicated that effects of symptom3 and symptom5 are underestimated, fever, labwbc2, and labneu2 are overestimated the effects, whereas rebten and movecough are not statistically significant.
- Thus, coefficients of these 5 varibles should be used to calibrate specific variables
- ***Calibrate specific predictors: symptom3 symptom5, fever2, labwbc2, labneu2
- gen score3 = (-3.374991+cf) + cal_b*(.7978508*symptom3 + 1.042774*symptom5 + .7787298*movecoug + 1.529419*rebten + 1.636311 *fever2 + .9095179*labwbc2 + .6898424*labneu2) + (1.284446*symptom3)+ (1.138048*symptom5)+ (-1.331718*fever2) + (-1.353248*labwbc2) + (-1.236354*labneu2)
- sum score*
- estat gof, gr(6)
- roctab appen score3

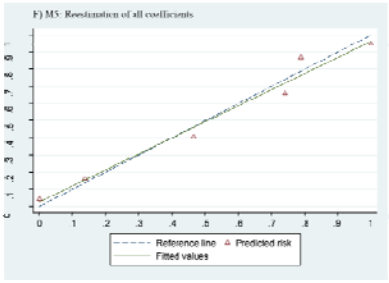


**SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS**
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

- M4: M2+ Stepwise selection of additional predictors.
- For this step, one or more predictors are not included in the original model, or
- new marker is needed to include in the model.
- Let's start with
- cal_b0 + cal_b*xb
- but we need to re-select what variables should be included in xb among 7 variables
- sw, pr(.05): logit append symptom3 symptom5 move fever2 rebten labwbc2 labneu2
- begin with full model
- only 3, i.e., symptom3, symtom5, rebten are ramained in the model!!
- gen score4 = (-3.374991-cf) + cal_b*(.7978508*symptom3 + 1.042774*symptom5 + .7787298*movecoug + 1.529419*rebten + 1.636311 *fever2 + .9095179*labwbc2 + .6898424*labneu2) + (1.83606)*symptom3 + (1.768313)*symptom5 + (1.816649)*rebten
- logit append score4

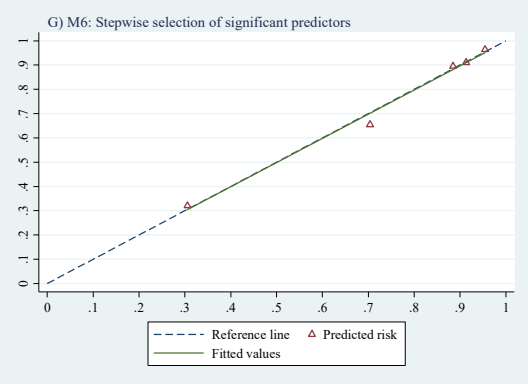
**SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS**
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

- Re-estimate all coefficients using validation data only
- logit append symptom3 symptom5 move fever2 rebten labwbc2 labneu2
- roctab append score5
- estat gof, gr(6) tab
- *** Prepare p data
- disp chiprob(4, 5.177072)
- ***Calibration plot
- lab var p5 ***Predicted risk***
- twoway (line obs_p5 ref p5 ref, sort lpattern(dash)) (scatter p5 obs_p5, sort msymbol(triangle_hollow) ylabel(0(.1)1) xlab(0(.1)1)) (lfit p5 obs_p5)
- ci mean o e



SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

- Model 6: M5 + stepwise selection; one or more predictors are not included
- sw, pr(.05): logit append symptom3 symptom5 move fever2 rebten labwbc2 labneu2
- begin with full model
- estat gof, gr(6) tab
- ***Calibration for m6
- ***Prepare p data
- disp chiprob(3,3.257358)




SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

- **We propose to choose the predictive score according to the prevalence of appendicitis as.**

**Prevalence of appendicitis**

**Less than 49%**

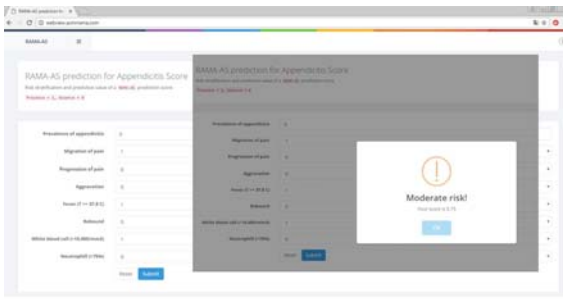
y= -3.628+0.915(0.77M+1.03P+0.77A+1.62F+1.52R+0.95F+0.81N)+1.84M+1.77P+1.82R

**49 to 77% or unknown**

y= -3.40+0.77M+1.03P+0.77A+1.62F+1.52R+0.95W+0.81N

**More than 77%**

y= -2.82+0.78(0.77M+1.03P+0.77A+1.62F+1.52R+0.95W+0.81N)+0.96M+1.13P+2.62R



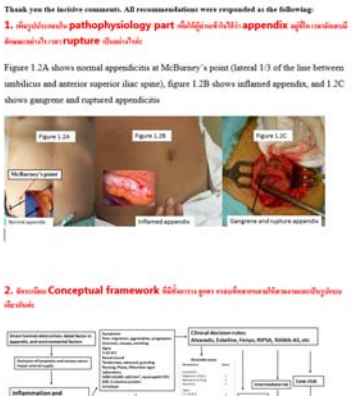

M=migration of pain, P=progression of pain, A=aggravation of pain by cough or movement, F=fever, R=rebound tenderness, W= white blood cell count, N=neutrophil

Although the RAMA–AS did not perform well in the external data when compared to the derive data, it still could well discriminated appendicitis from non appendicitis in secondary care setting (Chaiyaphum Hospital) and School of Medicine setting (Thammasat Hospital).

SECTION FOR CLINICAL EPIDEMIOLOGY & BIOSTATISTICS
Faculty of Medicine, Ramathibodi Hospital, Mahidol University

# PITFALL VS SERENDIPITY

- No success without pitfalls
- It is not a trap
- It is a serendipity
- Make greater success without concurrence


Thank you the incisive comments. All recommendations were responded as the following:

1. [illegible] pathophysiology part [illegible] appendix [illegible] rupture [illegible]

Figure 1.2A shows normal appendicitis at McBurney's point (lateral 1/3 of the line between umbilicus and anterior superior iliac spine), figure 1.2B shows inflamed appendix, and 1.2C shows gangrene and ruptured appendicitis



2. [illegible] Conceptual framework [illegible]



Ammarin Thakkinstian

Inbox

Hi Chumpon,
Here are my changes for Chapter 3..
Will work on the rests and give you comments soon..
Cheers,
A

**From:** Chumpon Wilasrusmee
**Sent:** Saturday, September 24, 2016 5:28 PM
**To:** Pat Woratanarat
**Cc:** ptomtitchong@gmail.com ; Panuwat Lertsithichai ; Amm
stephenramathibodi@gmail.com ; Napaphat Proprom
**Subject:** RE:

เรียนอาจารย์ทุกท่านผมขอส่ง files ที่แก้ไขครับ

Pat Woratana... เรียนอาจารย์จุมพลที่นับถือ เรียบร้อย

Vijj Kasemsup

**To:** Pat Woratanarat [pataraw@yahoo.com]; Chumpon Wilasrus
**Cc:** ptomtitchong@gmail.com; Panuwat Lertsithichai; Ammarin

Inbox

เรียนทุกท่าน

ผมเห็นด้วยกับอ.ภัทรวัณย์ ไม่มีอะไรเพิ่มเติมครับ

ขอบคุณครับ

วิชช์