Qualifying Examination

for Doctor of Philosophy (Epidemiology and Biostatistics)

(Duration: 09.00 – 16.00; Total score: 100)

*Students can use books or internet. Communication with anybody else is strictly prohibited.*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Keywords for scoring

**Part A: Epidemiology** (40 scores)

**Based on the attached article, please answer the following questions:**

1) What is the main conclusion of this paper? (2 scores)

**Ans**: LBW has a tendency to cause high adult blood pressure in South Asian region.

2). What is the magnitude of effect that the author used for answering the primary research question? If there was none or inappropriately reported, please specify what should be reported for the research question? (8 scores)

**Ans**: The last section of the RESULTS indicated the authors emphasis in p-value for the interpretation. Nonetheless, the OR can be considered as either weak or inconclusive:

SBP (OR = 2.89; 95% CI: 1.01 to 8.25; P = 0.04)

DBP (OR = 0.75; 95% CI: 0.22 to 2.16; P = 0.62)

Hypertension (OR = 3.15; 95% CI: 1.17 to 9.35; P = 0.03)

Inappropriately reporting- here are what should be:

* Number of outcome must always be presented (i.e., %^SBP, %^DBP, %Hypertension) so that model fitting strategies can be examined.
* The logistic regression model is likely over fitting as indicated by the wide Cis.
* Screening for candidate variables based on p-value can be misleading as shown that smokers are double among the NBW as that of LBW.
* Baseline characteristics of the respondents should be compared to the non-respondents to allow assessing the effect of selection bias in a study with a very high rate of non-responses.
* The reported model equations are meaningless to this study and should be omitted.

3). What are biases that could make the results invalid? Please specify name and roles of each bias, i.e., the conclusion could be distorted in what way and by what mechanism. (10 scores)

**Ans**:

* Selection bias due to extremely large number of non-responses (15,000 -> 217 -> 122). Respondent of hospital survey is likely to have health conditions, hence, over-estimate the magnitude of effect.
* Confounding effect due to smoking and alcohol consumption-> NBW had two folds of the LBW causing bias toward the null or under-estimate the magnitude of effect.

4). Will you recommend this paper as a piece of evidence for health care provider? Please describe. (10 scores)

**Ans**: No.

* Weak and inconclusive evidence
* Small size of sample under a large non-responses study population
* Not accounted for known confounders such as smoking and alcohol consumption

5). Please summarize in one paragraph of less than 10 lines the main points that will improve the quality of this paper- pretending you are writing a letter to the editor of this journal regarding this paper. (10 scores)

**Ans**:

* Number of outcome must always be presented (i.e., %^SBP, %^DBP, %Hypertension) so that model fitting strategies can be examined.
* The logistic regression model is likely over fitting as indicated by the wide Cis.
* Screening for candidate variables based on p-value can be misleading as shown that smokers are double among the NBW as that of LBW.
* Baseline characteristics of the respondents should be compared to the non-respondents to allow assessing the effect of selection bias.
* Model equation is meaningless to this study and should be deleted.
* Categorization of continuous variables leads to a less efficient in this study. The IDV (birth weight) and the DV (blood pressure) can be analyzed as continuous to investigate whether or not birth weight had an effect on the blood pressure, hence, multiple linear regression can be more appropriate.

**Part B: Biostatistics** (30 scores)

**Based on the data set with a description attached at the end of this material, please answer the following questions:**

6). Plan for data analysis: Please provide a dummy table, the table without number being presented, that is the main finding to answer the research question for this study. (10 scores)

**Ans**:

Treatment effect for each time point after treatment and the overall effect

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estrogen | Placebo | Different | 95%CI | p-value |
| Baseline | x.x | x.x | N/A | N/A | N/A |
| Month 1\* | x.x | x.x | x.x | x.x to x.x | 0.xxx |
| Month 3\* | x.x | x.x | x.x | x.x to x.x | 0.xxx |
| Month 6\* | x.x | x.x | x.x | x.x to x.x | 0.xxx |
| Overall\*\* | x.x | x.x | x.x | x.x to x.x | 0.xxx |

\* Treatment effect for each time point: The confidence interval (CI) and p-value were adjusted for baseline depression score using analysis of covariance (ANCOVA)

\*\* The overall effect combing results after the treatment and adjusted for baseline depression score as well as accounted for clustering effect due to repeated measurement using generalized estimating equations (GEEs)

7). Please provide Stata or R commands (Stata do file or R scripts) to obtain the results for the dummy table you’ve provided. (10 scores)

**Ans**:

clear

set more off

gen grp = 1

replace grp = 2 if group == "B"

gen bmi = weight / ((height/100)^2)

\*Table 1: Baseline comparison

tabstat age, stat(n mean sd median min max) by(grp)

tabstat bmi, stat(n mean sd median min max) by(grp)

tabstat pdepres, stat(n mean sd median min max) by(grp)

\*Table 2: Main findings

\*Effect for each time point using ANCOVA

regress depres1 grp pdepres

regress depres3 grp pdepres

regress depres6 grp pdepres

\*Overall effect using GEE

reshape long depres, i(id) j(month)

xtset id month

xtgee depres grp pdepres, link(identity) fam(gaussian) corr(exchangeable) robust

8). Please summarize results as an abstract. Note that this must be presented with real results obtained from your own analysis. (10 scores)

**Ans**:

|  |
| --- |
| **Background**: Depression affects mental health among women in their estrogen-producing years more often than men or postmenopausal women. Estrogen is also linked to mood disruptions that occur only in women - premenstrual syndrome, premenstrual dysphoric disorder, and postpartum depression. Information on the efficacy of estrogen oil capsule in treating postnatal depression remained limited.    **Objective:**  To investigate efficacy of estrogen oil capsule in treating postnatal depression in Thai women.  **Methods:** This is a randomized controlled trial conducted in a teaching hospital in northeast of Thailand. Thai women who met screening criteria were randomly assigned to either a placebo control group (group B, n=25) or estrogen oil capsule group (group A, n=24). Prior to the first treatment, all patients took the Edinburgh Postnatal Depression Scale (EPDS). The EPDS data was collected monthly for six months once the treatment began. Higher scores on the EDPS are indicative of higher levels of depression.  **Results**: The two experimental groups were similar at baseline regarding age, body mass index, and EPDs before the treatment. Estrogen group had a significantly lower EPDs than placebo group at month 1 (Mean difference EPDs: x.x; 95%CI: xx to xx, *P* = 0.xxx), at month 3 (Mean difference EPDs: x.x; 95%CI: xx to xx, *P* = 0.xxx), and at month 6 (Mean difference EPDs: x.x; 95%CI: xx to xx, *P* = 0.xxx). Overall, estrogen group remain having a significantly lower EPDs at month 1 (Mean difference EPDs: 4.0; 95%CI: 1.9 to 6.0, *P* < 0.001), or equivalently, 20% reduction in reference to the baseline EPDs.  **Conclusions**: Estrogen oil capsule can reduce postnatal depression. |

**Part C: Dissertation in brief for all students except Ms. Worawan** (30 scores)

**Based on your own your plan on your dissertation, please answer the following questions:**

9) Summarize your dissertation in one paragraph of no more than 300 words (5 scores)

10). Describe all possible alternative statistical methods that could be efficiently applied to your main paper as well as their advantages and disadvantages and justify why the one you selected is the best choice. (10 scores)

11) What is the main weakness of your thesis and how it affects the main conclusion? (5 scores)

12) What is/are the impact(s) that would be expected from the findings of your thesis? (10 scores)

**Part C: Dissertation (ONLY For Ms. Worrawan Poochada)** (30 scores)

**Please answer the following questions in brief for this written qualifying exam and later prepare a presentation (15-20 min) of your answers for oral qualifying exam**

1. Based on the study result from research and the surveillance report from health data, which one is more suitable to describe situations of work-related diseases among Thai farmers? What is your reason regarding strength and weakness to support the answer? (5 scores)
2. Have you ever found misrepresented data or distortion of report in the morbidity rate of occupational diseases among Thai farmers? If so, please describe in list of possible causes. (5 scores)
3. What is the solution for problems of distorted report and how to develop methods for the best representative of health data to accurate, accessible and actionable information of occupational diseases among Thai farmers? Please describe your idea in brief. (5 scores)
4. For your PhD dissertation in program of Epidemiology and Biostatistics, please summarize your concept paper in 3-4 paragraphs for the rational explanation of your research concerning occupational diseases in Thai farmers, the methodology (design, data collecting and analysis), and the impacts that would be expected from the findings. (15 scores)

GOOD LUCK