Abstract

Rationale: Most of statistical methods used in meta-analysis assume individual subjects as units of randomization. Meta-analyses involving cluster randomized trials may lead to additional sources of heterogeneity beyond those elevated by meta-analyses involving only individually randomized trials. The appropriate statistical analysis to these meta-analyses must take into account potential heterogeneity in the cluster randomized trials. A substantial amount of literature covering statistical methodologies used in meta analyses can now be found. Most of them, however, assume individual subjects as units of randomization. Therefore, there may remain some questions that need to be investigated in the area of meta analyses related to the inclusion of cluster randomized trials.

The general linear mixed model (GLM) has been proposed to explain heterogeneity in meta-analysis where the treatment effect is measured in binary outcome. Log-relative measure is used as a response variable. The parameter estimation is based on assumption of normal distribution of random effects. The generalized linear mixed model (GLMM) under unspecified distribution of random effects may be an alternative choice. The two approaches allow the inclusion of some covariates of trial level and subject level. Therefore it is interesting to explore potential of the two approaches in meta-analysis involving cluster randomized trials in binary outcome.

Objective: Two potential non-Bayesian approaches of GLM and GLMM are explored to identify and explain heterogeneity in meta-analyses involving cluster randomized trials comparing two treatment groups measured in binary outcome.

Methods: The two approaches of GLM and GLMM are studied and evaluated their potential in term of methodological aspects, results provided, strengths and limitations of these approaches and exemplified in three published meta-analyses involving cluster randomized trials. The first meta-analysis includes eight community-based trials. They were performed in developing countries to examine the relationship of vitamin A supplementation and mortality in children aged 6 to 72 months. None of the trials assigned individual children to treatment groups. The second meta-analysis comprises fewer trials of 8, which is performed to evaluate the effect of mammographic screening on reduction of breast cancer mortality. The third meta-analysis is done to assess the effectiveness of multiple risk factor interventions to reduce cardiovascular risk factors from coronary heart disease. Analysis is performed in the 14 trials included that provided smoking prevalence outcome. For each meta-analysis, observed log-relative risks for individual trials are fitted to the GLM as a continuous response. The trials included are classified to two categories according to randomization units, clusters and individually, and called randomization design variable. This variable is treated as a covariate of the model. The model parameters are estimated with the restricted maximum likelihood (REML) under the normality assumption of random effects via MLwiN software. For the GLMM, observed frequencies of the outcome for each treatment group are used rather than the observed log-relative risks for individual trials. A canonical link function of the observed mean proportions is associated with linear predictors model of which treatment and randomization design are treated as covariates. Here, the treatment effect can be treated as random treatment effects. The maximum likelihood estimates of the model parameters are obtained non-parametrically under a discrete mixture distribution of random effects for K components, which is implemented by the EM-algorithm procedure via S-plus software. Maximum posterior probability is used to classified trials to each component.

Results: The two approaches shown that the covariates effects and variability of random effects from the models easily explained heterogeneity between trials. Results of numerical

examples are presented in topic 6 and 7. The GLMM is superior to the GLM in some aspects. The GLMM gives further heterogeneity information from random treatment effects. In addition, the approach provides component (or subgroup)-specific treatment effect and trial classification according to the optimal components. This is very useful in further explaining the heterogeneity that might be beyond the effects found in the model.

Conclusions: The GLMM approach provides more information for explaining heterogeneity effect in meta-analyses involving cluster randomized trials. However, care should be taken when interpreting the covariates effects of the model because inference on these effects obtained from a discrete mixing distribution have not been ruled out. Nevertheless, the GLMM would be much more efficient when it is applied to large meta-analyses.

บทคัดย่อ

ความเป็นมา : วิธีการวิเคราะห์หน่วยเดียวกับทางสถิติที่ใช้ใน meta–analysis ส่วนใหญ่เป็นวิธีการที่ใช้กับข้อมูล ซึ่งมาจากหน่วยศึกษาที่เป็นหน่วยเดียวกับหน่วยสุ่ม meta-analysis ที่เกี่ยวข้องกับผลลัพธ์จากการทดลองที่มีกลุ่ม (cluster) ของหน่วยศึกษา (individual units) เป็นหน่วยสุ่มรับสิ่งทดลองอาจทำให้เกิดความแตกต่าง (heterogeneity) ที่นอกเหนือจาก meta-analysis ที่มีหน่วยศึกษาเป็นหน่วยเดียวกับหน่วยสุ่มรับส่งทดลอง วิธีการทางสถิติที่เหมาะสมกับ meta-analysis ที่เกี่ยวข้องผลลัพธ์จากการทดลองที่มี clusters ของหน่วยศึกษาเป็นหน่วยสุ่มรับสิ่งทดลอง จะต้องพิจารณาอิทธิพลที่เกิดจาก clusters ด้วย

ได้มีการเสนอ General linear mixed model (GLM) มาวิเคราะห์ข้อมูลของผลลัพธ์สิ่งทดลองที่วัดออกมาเป็นข้อมูล binary โดยใช้ค่า log-relative measure เป็นตัวแปรตามของโมเดล การประมาณค่าพารามิเตอร์จะขึ้นอยู่กับการแจกแจงแบบปกติของ random effects Generalize linear mixed model (GLMM) เป็น model ที่ไม่ต้องการรูปแบบการแจกแจงของ random effects ในการประมาณพารามิเตอร์ GLMM อาจเป็นอีกวิธีหนึ่งที่นำมาใช้ได้ GLM และ GLMM สามารถให้ข้อมูลอิทธิพลของ covariates ทั้งที่ระดับสิ่งทดลองและระดับหน่วยศึกษา ดังนั้นการศึกษาความสามารถของวิธี GLM และ GLMM ในการวิเคราะห์ข้อมูลของ meta-analysis ที่เกี่ยวข้องกับผลลัพธ์จากการทดลองที่มีกลุ่มของหน่วยศึกษาเป็นหน่วยสุ่มรับส่งทดลอง จึงเป็นสิ่งที่น่าสนใจ

วัตถุประสงค์ : เพื่อศึกษาความสามารถของ GLM และ GLMM ในการตรวจสอบและอธิบายความแตกต่างใน meta-analysis ที่เกี่ยวข้องกับผลลัพธ์จากการทดลองที่มีกลุ่มของหน่วยศึกษาเป็นหน่วยสุ่มรับส่งทดลอง ซึ่งวัดประสิทธิภาพสิ่งทดลองเป็น binary data

วิธีการศึกษา : ในการศึกษาครั้งนี้ ได้ทำการประเมินความสามารถของ GLM และ GLMM ในประเด็นเกี่ยวกับ วิธีการ ผลลัพธ์ ที่ได้จากแต่ละวิธี ข้อดีและข้อจำกัดของแต่ละวิธี การใช้ GLM และ GLMM ในการวิเคราะห์ข้อมูลจริงจาก meta-analysis ดังกล่าว 3 เรื่อง ได้แก่เรื่องแรก meta-analysis ที่ศึกษาเพื่อหาความสัมพันธ์ระหว่างวิตามินเอ และการตายในเด็กอายุ 6 ถึง 72 เดือน โดยรวบรวมข้อมูลจาก 8 การทดลองที่ศึกษาในชุมชนของประเทศกำลังพัฒนา และในทุกการทดลองไม่มีการสุ่มเด็กรับวิตามินเอ เรื่องที่สองเป็น meta-analysis ของข้อมูลการทดลอง 8 เรื่อง ที่ศึกษาประสิทธิภาพของ memmographic screening ในการลดการตายของมะเร็งเต้านม เรื่องที่สามเป็นเรื่อง meta-analysis ของข้อมูลการทดลอง 14 เรื่อง ที่ศึกษาประสิทธิภาพของ multiple risk factor interventions ในการลดปัจจัยเสี่ยงของโรคหัวใจโคโรนารี่ โดยมีอัตราชุกของการสูบบุหรี่เป็นผลลัพธ์ที่ใช้วิเคราะห์ การวิเคราะห์ด้วย GLM ใน meta-analysis แต่ละเรื่อง ค่าสังเกต log-relative risk ของแต่ละการทดลองจะนำมาใช้เป็นตัวแปรตามแบบต่อเนื่องในโมเดล ลักษณะของหน่วยสุ่มรับสิ่งทดลองนำมาศึกษาเป็นตัวแปร covariates ที่แยกออกเป็น 2 กลุ่มตามลักษณะของหน่วยสุ่มรับสิ่งทดลอง ซึ่งอาจจะเป็นแต่ละหน่วยศึกษา (individual) หรือ กลุ่มของ

หน่วยศึกษา (cluster) การประมาณพารามิเตอร์ของโมเดลใช้ restricted maximum likelihood ภายใต้ข้อสมมุติของการแจกแจงแบบปกติของ random effects โดยวิเคราะห์ด้วยโปรแกรม ML win

สำหรับการวิเคราะห์ข้อมูลดังกล่าวด้วย GLMM ค่าสังเกตจำนวนของผลลัพธ์ที่เกิดขึ้นในแต่ละกลุ่มของสิ่งทดลองของการทดลองแต่ละเรื่อง คือ ข้อมูลของตัวแปรตาม โมเดล canonical link function ของค่าสังเกตค่าเฉลี่ยสัดส่วนจะสร้างจากโมเดลตัวประมาณค่าเชิงเส้นตรงที่มีสิ่งทดลองและแบบการสุ่ม (randomization design) เป็นตัวแปร covariates ในการวิเคราะห์ของ GLMM สามารถกำหนดให้อิทธิพลของสิ่งทดลองเป็นแบบสุ่มได้ การประมาณค่าพารามิเตอร์ของโมเดลใช้วิธี non-parametric maximum likelihood ภายใต้การแจกแจงของ random effects แบบไม่ต่อเนื่อง ที่แยกออกเป็น k components ซึ่งทำการวิเคราะห์โดยใช้ EM-algorithm ด้วยโปรแกรม S-plus Maximum posterior probability ที่ได้จากการวิเคราะห์นำมาใช้ในการจัดกลุ่มการทดลองให้กับแต่ละ components

ผลลัพธ์ของการศึกษา : จากการวิเคราะห์ด้วย GLM และ GLMM แสดงให้เห็นว่า อิทธิพลของ covariates และความแตกต่างของ random effects สามารถใช้อธิบายความแตกต่างระหว่างการทดลองได้อย่างง่ายดาย ผลลัพธ์ในเชิงตัวเลขที่ได้จากการวิเคราะห์ในข้อมูลตัวอย่าง meta-analysis ได้แสดงไว้ในหัวข้อที่ 6 และ 7 ของรายงานฉบับนี้ และพบว่าวิธีของ GLMM ดีกว่า GLM ในบางประเด็นซึ่งได้แก่ ความสามารถในการวิเคราะห์อิทธิพลของสิ่งทดลองเชิงสุ่ม ซึ่งนำมาใช้อธิบายข้อมูลความแตกต่างระหว่างการทดลองได้ นอกจากนี้ GLMM ยังให้ข้อมูลเกี่ยวกับอิทธิพลของสิ่งทดลองเฉพาะในแต่ละ component ซึ่งนำมาใช้ในการแยกการทดลองเป็นกลุ่มๆ ตามจำนวนของ optimal components ข้อมูลเหล่านี้เป็นประโยชน์อย่างยิ่งในการอธิบายความแตกต่างระหว่างการทดลองที่เกิดขึ้น ซึ่งเป็นอิทธิพลที่นอกเหนือขอบเขตของข้อมูลที่อธิบายได้จากโมเดล

สรุปผลการศึกษา : GLMM เป็นวิธีการวิเคราะห์ที่สามารถใช้ข้อมูลเพื่ออธิบายความแตกต่างที่เกิดขึ้นใน meta-analysis ที่เกี่ยวข้องกับการทดลองที่มีกลุ่มของหน่วยศึกษาเป็นหน่วยสุ่มรับสิ่งทดลอง อย่างไรก็ตามควรต้องระมัดระวังในการแปลผลอิทธิพลของ covariates เนื่องจากยังไม่มีหลักฐานการตรวจสอบความถูกต้องของการอนุมานอิทธิพลดังกล่าวที่วิเคราะห์ได้จาก discrete mixing distribution แม้จะด้วยเหตุผลดังกล่าว แต่พบว่าการใช้ GLMM ใน meta-analysis ที่รวบรวมจำนวนการทดลองมาศึกษาเป็นจำนวนมากจะให้ผลลัพธ์ที่มีประสิทธิภาพดีขึ้น