Opportunities and Challenges of an International Collaborative Pooled Analysis of Dietary Cohort Studies

Jung Eun Lee
Department of Food and Nutrition, Seoul National University, Seoul, Korea

Summary
International collaborative work of prospective cohort studies has been conducted for a few decades. It has provided a rich source for researchers worldwide to investigate the role of dietary factors in the development of many diseases. A pooled analysis of the existing cohort studies has strengths, including increased statistical power, increased generalizability, the disclosure of unpublished work, use of the standardized approaches, a wide range of dietary exposures, inclusion of diverse populations, and a build-up of partnerships of investigators. The advantages of pooled analyses are significant in nutritional epidemiology. An increase in statistical power and analysis of multilevel exposures in the consortium data allow us to have better insight and knowledge of dietary factors related to chronic diseases. However, facilitating data harmonization of foods, food groups, and nutrients in the consortium data remains challenging. Combining and standardizing dietary data is complex, given differences in food frequency questionnaires, foods assessed, dietary exposure categories and intake levels, the validity of the dietary assessment, and nutrient database availability across individual studies. Although there are limitations in the pooled analysis, international collaborative pooled analysis is a valuable and efficient means to produce important scientific evidence. In addition, consortial research needs to continue to expand in the era of moving toward a diverse world and precision medicine.

Key Words
pooled analysis, nutritional epidemiology, cohort consortium

Collaborative studies that combine individual cohort data have become a common approach (1). The Pooling Project of Prospective Studies of Diet and Cancer was founded in 1991 (2) and provided important scientific evidence. The National Cancer Institute (NCI) cohort consortium began in 2001 upon agreement on the value of pooled analysis (1). The Asia Cohort Consortium was proposed in 2004 (3), allowing the expansion of epidemiologic findings to Asian populations, which have been somewhat underrepresented in the epidemiologic evidence. In particular, investigators who are interested in dietary roles in chronic disease development have been motivated to establish a collaborative work group because of the need for a larger sample size. This could be due to the following several reasons: 1) Dietary factors account for the long latency period of disease development. Therefore, a prolonged observation period is necessary to investigate the effect of diet on chronic disease etiology; 2) the effect of dietary factors is relatively small compared to some environmental risk factors, including smoking, asbestos, and toxic agents; 3) measurement error may be inherent in the assessment of dietary intake; 4) there is heterogeneity even within the same disease, and therefore, subgroup analysis according to subtypes of disorders is an ideal strategy to explore the association; and 5) precision medicine has drawn attention and involves subgroup analyses based on a multidimensional dataset, including variables of diet, biomarkers, genetic factors, diagnostic and clinical factors, and microbiome. Underpowered findings in diet and chronic disease research partly have often discouraged investigators from spending time and energy publishing the data.

Combining the existing cohort data becomes appealing because starting a new large-scale cohort requires substantial resources, such as funding, time, and human efforts. Additionally, it is possible that the new cohort remains limited because cohorts tend to have a limited range of ethnic groups, exposures, and age groups. International collaborative work of pooled analysis of existing cohort data can efficiently provide evidence covering diverse ethnic and age populations and a wide range of exposures.

Opportunities for an international collaborative pooled analysis of dietary cohort studies

Pooled analysis of individual-level cohort data can increase statistical power. An increase in sample size enables investigators to examine rare outcomes and the associations between diet and health outcomes according to various subgroups. Studying the effect of diet on health outcomes across genetic, molecular, environmental, clinical, socioeconomic, and psychological factors has become a common approach to addressing personalized prevention and treatment. Assembly of individual data can improve the generalizability because the pooled analysis can encompass diverse ethnic and age populations. Examining a wide range of dietary exposures is an attractive strength because the possible reason for null findings in nutritional epidemiologic studies

E-mail: jungelee@snu.ac.kr
can be a relatively narrow range of dietary exposure of interest in the study population. For example, if the association with disease is limited to low intake, comparing high intake with medium intake would not show an association in the populations with relatively high exposure intakes. However, a pooled analysis of individual cohorts whose exposure levels range from low to high can provide better insight into the shape of the association. Because the assembly of individual data incorporates broad diversity, it is possible to examine the potential heterogeneity in the association between dietary factors and health outcomes. Additionally, a pooled analysis has opportunities to present unpublished findings of individual cohort studies, which did not analyze or present the data partly due to limited statistical power. Disclosure of unpublished data can lead to avoiding publication bias. Moreover, a significant advantage includes training of junior investigators through senior investigators’ constructive feedback and a build-up of international collaborative partnerships. Under international collaboration work, this can lead to the advance of cohort data and the development of new cohorts in underrepresented populations.

### Challenges of an international collaborative pooled analysis of dietary cohort studies

Heterogeneity sources related to dietary research include differences in dietary assessment tools, food items included, dietary exposure categories, intake levels, the validity of the dietary assessment, and nutrient database availability across individual studies. The lead investigator needs to examine whether these differences can modify the findings. The numbers of food items assessed in the dietary assessment instruments are different across studies. For example, the number of fruit and vegetable items in each study ranged from 9 to 54 in the fruit and vegetable intake and lung cancer of the Pooling Project of Prospective Studies of Diet and Cancer (4). Food items in one food group are not uniform across studies. For example, fermented salted vegetables can be included in the vegetable group in East Asian cohort studies but not in other cohort studies. Additionally, eating behaviors and cooking styles for specific food intake are diverse. For example, people in some countries often add sweetened additives to their beverages, such as condensed milk when they drink tea. Dietary exposures are categorized differently, and intake levels vary across studies. The validity levels of dietary assessment tools and nutrient database availability vary across studies. Therefore, lead investigators may need to collect the validation study findings and review the validity levels for dietary factors of interest. When the analysis is conducted at the nutrient level, the nutrient database is vital information. Given these differences, the harmonization of dietary data is challenging. Investigators can harmonize the foods or request each cohort to generate dietary variables. Either way, investigators need to be involved in the decision-making process of harmonization.

Another major challenge is the operation of the collaborative platform. The consortial team should employ the operating system and infrastructure before assembling the individual data. Consortia administrative work often requires considerable paperwork, a multilayer decision-making process, time-consuming communications, and extensive discussion.

### Approaches to overcome challenges

A coordinating center is needed to operate the consortium, gather the data, and harmonize and analyze the data. The setup of a secure portal and portal server hosting services are essential steps. A centralized data repository and remote analysis have become more available due to advances in information and communication technology.

Annual meetings and satellite working group meetings are important to keep investigators informed, share the findings and challenges, and discuss plans and concerns. The working group meeting is essential for dietary research because the leading investigator often lacks sufficient knowledge of each country’s foods and eating behaviors. Additionally, because the study population characteristics according to specific food intake differ across the studies, potential confounding factors may not be homogeneous. The working group meetings should be able to integrate critical feedback and knowledge. Because the heterogeneity across dietary cohort studies can be substantial, examining the heterogeneity source and conducting a sensitivity analysis are necessary. Food items included in the food group also need to be listed and compared across studies. Each lead investigator needs to understand what each food is (including knowing the original name of the food), request information about the food from coinvestigators from that

---

**Table 1. Opportunities and challenges of an international collaborative pooled analysis of dietary cohort studies.**

<table>
<thead>
<tr>
<th>Opportunities</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Increase in sample size</td>
<td>- Heterogeneity across studies</td>
</tr>
<tr>
<td>- Increase in generalizability</td>
<td>- Harmonization of dietary variables</td>
</tr>
<tr>
<td>- Examination of rare outcomes</td>
<td>- Set-up of operating systems and infrastructures (e.g., secure portal, steering committee, and governance rules)</td>
</tr>
<tr>
<td>- Examination of a wide range of exposure</td>
<td>- Considerable paperwork (e.g., data use agreement form, participating agreement form for subproject, code book and study dictionary of each cohort)</td>
</tr>
<tr>
<td>- Assessment of the heterogeneity across population subgroups</td>
<td>- Multilayer decision-making process</td>
</tr>
<tr>
<td>- Release of unpublished data</td>
<td>- Extensive discussion for an agreement</td>
</tr>
<tr>
<td>- Provision of scientific knowledge and constructive feedback</td>
<td>- Intensive and lengthy data analysis process</td>
</tr>
<tr>
<td>- Building up of international collaborative partnerships</td>
<td></td>
</tr>
<tr>
<td>- Incubation of new cohorts</td>
<td></td>
</tr>
</tbody>
</table>

---

Lee JE
country, and discuss how each study groups the food items. Additionally, investigators should keep in mind that eating behaviors could be different even for the same food item.

It is suggested that the pooled analysis of dietary studies may need to estimate the parameters within individual-level data and then combine the parameters. Additionally, examining the heterogeneity and heterogeneity sources are essential steps. Lead investigators can conduct various analyses by treating dietary exposures as continuous variables, quantiles, and absolute cutoff categories. It is important to address the associations using different approaches and check the robustness of the findings. When dietary information is collected through semiquantitative or quantitative food frequency questionnaires, the frequency per day of each food item can be calculated based on median values. Therefore, quantifying the amount of food intake is feasible, although the frequency categories and portion sizes are different across studies. However, when investigators use categories instead of continuous variables, the categories tend to be regrouped into fewer categories.

Conclusions

International collaborative pooled analysis of dietary cohort studies provides an opportunity to examine the hypothesis of dietary factors and various health outcomes with great statistical power. Investigators can standardize dietary variables and analyze the heterogeneity in the associations for dietary factors across subpopulations and subendpoints with a wide range of intakes. These strengths allow investigators to study the association across the distribution from low to high intakes and release unpublished data. The collaborative framework can also build international partnerships and knowledge flow. Although there are limitations in the pooled analysis, international collaborative pooled analysis is a valuable and efficient means to produce important scientific evidence. In addition, consortial research needs to continue to expand in the era of moving toward a diverse world and precision medicine.

Disclosure of state of COI

No conflicts of interest to be declared.

Acknowledgments

The author would like to thank the researchers and staff of the Asia Cohort Consortium Coordinating Center.

REFERENCES


