【講演3】

竹内　文乃（独立行政法人 国立環境研究所 環境健康研究センター 研究員）

“Insights from a nationwide children’s study”

Thank you for inviting me and thank you for the introduction. Today, I’d like to talk about the ‘Insights from a Nationwide Children’s Study’. I belong to National Institute for Environmental Studies at Tsukuba, Ibaraki Prefecture and I’m in charge of biostatistics. First of all, I have to do the stereotype declaration about my presentation. Today’s talk is not governmental or our institutional official presentation, it’s my personal opinion.

And next, I introduce myself. I am biostatistician, belonging to the Biometric Society, and my major is epidemiology, analysis of epidemiological data. It means not intervention, just observational study. My statistical interest is Bayesian bias analysis with observational data.

Now, I work at the Office of Japan Environment and Children’s Study. We call this study, JECS. JECS is a birth cohort study. Cohort study is one of the famous studies design in epidemiology. I hope you know cohort study. In short explanation, birth cohort study recruit pregnant women and follow up the children until they become adult.

There are many kinds of big data, but it can be categorized into two types. One big data is collected automatically like an access log or electric medical record. Statisticians divide these kind of big data into signal and noise. On the other hand, our big data is not collected automatically. We need to design and prepare for beforehand data collection. So, it takes huge budget and time to collect data. For example, our study use almost 5 billion yen per year, it is same as Dr. Okubo’s presentation.

This is today’s outline. First, I will talk about why JECS Study is needed. And next, overview of JECS; study design, subject, aim and roadmap in the future. Some result and future statistical problem is presented at last. Some result of JECS is related to the psychological aspect of pregnant women.

I’m not sure about the definition of big data. Number of JECS subjects is more than 100,000 and we follow up our subjects from pregnant period to the child becomes 13 years. We have more than 2,000 variables from pregnant period to delivery period. So at the end of our study (at the time our subjects become 13yrs) number of variables must be millions I think. Is this the big data? Alright.

This slide shows what’s happening to today’s children. Congenital abnormalities have increased in Japan these days. Same as hypospadias, Down’s syndrome, immune system diseases, childhood asthma, endocrine and metabolic abnormalities such as children’s obesity. Birth defect have doubled over this 25 years, childhood asthma has tripled over this 20 years, and children’s obesity have increased by 150% over this 30 years. In Japan,

the age of pregnant women getting higher and higher, and it's also in United States. This must cause these defects and abnormalities, but it's not enough to explain this increase.

And this slide shows the rate of reproductive abnormalities and male birth ratio. Reduced ratio of male to female births, especially in metropolitan area like Tokyo, Osaka, and Kyushu, Fukuoka area, has dropped significantly in these 40 years. Some researches show that this tendency is remarkable in industrial areas, especially chemical industrial areas.

This slide shows the problem about psycho-neurodevelopment abnormalities. The left figure shows percentage of children receiving medical treatment for mental and behavioral disorders is increasing in Japan, and right figure is in the United States. The prevalence of children with autism increased and increased in these days.

There are many evidences and many hypotheses about the effect of environmental chemicals to children's psycho-neurodevelopment. Most famous and established evidence is that low level exposure of methyl mercury effect to childhood development, especially to their psycho-neurodevelopment, and low level exposure of lead effect to mental deterioration and decreased intelligence. Low level exposure of PCB and organic or authentic compounds effect to childhood development. These chemicals are not All of these, mercury, lead, PCB and these compounds affect the mental and psycho-neurodevelopment health effects is doubted.

There are more and more health problems threatening children's development and future health. Some environmental chemical compounds have evidence to cause carcinogen, low immunity, low IQ, low birth weight and so on. And we believe some unknown environmental chemicals, lifestyle and diet cause these health problems.

For example, the causal relationship between tobacco smoking and lung cancer is very clear, strong and specific relation. But the relationship between environmental chemicals and children's development is not so clear, not so strong and not so specific. So we need to conduct huge birth cohort to detect these kind of causal relationship with proper statistical power. So, this is why we need to conduct JECS Study with 100,000 subjects.

In these days, many developed countries reached the same conclusion as JECS. It means large birth cohort study is needed to protect future children's health. There are many birth cohort studies that have started in these days. More than 10 countries launch or prepare to conduct birth cohort study, which subject is more than 50,000, which shows as orange and green line at this slide. And more than 20 birth cohort study which subject is less than 10,000 shows as blue line at this slide.

I'd like to talk about the I4C framework at this slide. I4C stands for International Childhood Cancer Cohort Consortium. The reason why we need this kind of consortium is

that childhood cancer is one of the rare disease, the prevalence of children's cancer is less than 0.1%. In Japan we follow up 100,000 of children but only 200-300 children hood cancer can be detected. Sample size, 100,000 is not enough to detect causal relationship between environmental factors to children's cancer based on the statistical power's perspective. In future, the members of I4C gather and merge data sets to prevent children's cancer, so this is a big meta-analysis framework. JECS is the members of I4C, our study is also internationally important. This is our international outreach framework.

This slide shows our expected scientific output. Our expected scientific output classified into two aspects, one is direct results and the other is indirect results. The direct results identify environmental factors impacting children's health. This result helped to eliminate hazardous environmental chemicals. And next one is identifying gene relating chemical substance, diseases and disorders. These results help to prevent and reduce children's disease, and ensure good environment for children's development. The indirect results, we think, are to provide broad scope and to share research foundation for children's health, not limited to environmental factors. The second indirect results are to provide function as a biological specimen bank, bio-bank. And the third one is to provide function as a data archives and data-base. And the fourth one is most important to academic researcher, to cultivates or strengthens competencies of the nation's environment epidemiologic researchers. These are our expected scientific outputs.

I'd like to talk about overview of JECS Study from this slide. Our core hypothesis is chemical exposure during the fetal and in fetal stage, adversely affects children's health, and the goal of our study is to better understand the impact of environmental exposure on children's health and develop to improve their health and future environment. And environment includes indoor and outdoor air, water, soil, dust, diet, noise, socio-economic status, education, occupation, mental and social support and so on. Environmental factor is very broad concept, not only chemicals. And genetic factors also will be examined. Metabolic and other-omics data; genomics, epigenomics, and metabolomics will be also examined.

We expect four anticipated results. First one is to identify environmental factors impacting children's health. The second one is to develop risk-management system and address vulnerable in children. The third one is to ensure sound environment where the future generations are able to grow up in good health. The last one is to enhance global competitiveness and benefit Japan's nation interest. This is our overview of JECS.

Our study method is multi-layered prospective birth cohort study. This is our original concept, not technical term. 'Multi-layered' means that we collaborate with multi-level of center, such as local clinic, university hospital and data center. Birth cohort is technical

term. Our study is composed into three study. Main study, sub study and adjunct study. These are the three main components in our study.

Before JECS started, we conduct pilot study. This slide shows outline of pilot study. Pilot study was conducted beforehand main study to evaluate the feasibility, acceptability and cost. Whether everything is acceptable or not, we improve out information documents, questionnaires, study procedures and administration protocol. This is the pilot study. Study subjects in pilot study are completely different from main study.

And main study contain 100,000 mother-child pairs, and fathers from 15 study locations across Japan. We follow up participants until children become 13 yrs. Our recruit period was from Jan 2011 to March 2014.

Sub-study is a sub-cohort study of JECS. Sub-cohort contains 5,000 randomly selected subjects from main study population, which contains 100,000 subjects. The aim of Sub-study is to examine extended exposure outcome relationship. It means, we visit participant's home and conduct intensive evaluation. For example we will measure indoor exposure, outcome air, indoor, outdoor air, and interview study to detect children's developmental disorder. On the other hand, main study subject follow up only with questionnaire, but the sub-study subject follow up more intensively.

This is the third one, adjunct study. Adjunct study is conducted on the main study, which study protocols are reviewed and approved by the steering committee. Any member of JECS can propose adjunct study to examine their originally interested hypotheses. But they need to conduct adjunct study with budget which is independent from JECS.

This slide shows the types of epidemiological study design. Cohort study, case-control study, case cohort study, and nested case-control study. Case-control study is just for rare disease. Our study adopt cohort study and cohort based case cohort and nested case-control study. The pros of case cohort study and nested case-control study is to save bio-specimen. The amount of bio-specimen we collected and cost to measure chemicals are both limited. So we choose the Case-Cohort study or Nested Case-Control study to limit subject who is chosen and measure bio-specimen efficiently. In these study design, subjects who do not have any diseases are randomly selected as control, compared and detect a relationship between exposure and disease. JECS is flexible and developing study learned from other studies.

This is the summary table about our three types of studies, main study, sub study and adjunct study. Main study is 100,000 people follow up with questionnaire, and sub-study is intensive follow up, random sample from main study, 1,000 to 10,000 but it is now decided to 5,000 randomly sampled subjects from main study. And we prepare to accept many

types of adjunct studies.

We need to collaborate with any other academic researchers, industry and other governmental sections. Because we use 5 billion yen per year so we have many responsible for commit industrial area, academic findings, and Ministry of Environment (MOE). Whole budget of Ministry of Environment, whole budget is 2,000 billion yen per year so 5 billion is how big percentage of MOE's budget we consume.

This slide shows the concrete hypotheses of JECS. The left figure shows list of candidate exposure, and the right figure shows health outcome. The exposure is environmental factors, and chemicals, genetics, socioeconomic status and lifestyle. And health outcomes are physical development, congenital abnormalities, sexual differentiation abnormalities and psycho-neurodevelopment, compromised immune system and endocrine and metabolic abnormalities. So, there are more than 100 hypotheses, the relation between exposure and outcomes, so we need to detect which is sufficient evidence, and which is not so sufficient evidence.

And this slide shows why we need 100,000 study subjects. This is biostatistician's work to calculate sample size. This slide shows the sample size calculation table. We assume the simple 2x2 table, high exposure and low exposure, disease or not diseased, and conduct chi-square test. Significance level is 5% and the power is 80%. Of course, obesity – this is not disease, but popular, 10% prevalence and few percent, we can detect at 100,000 of children. But we cannot detect Down syndrome and hypospadias, and type 1 diabetes, we cannot detect in our JECS studies, so we have to do meta-analysis about these diseases to other countries, so we plan to, how we can do meta-analysis about these kinds of diseases, so this is a sample size table.

This is implementation framework. MOE is responsible for budgeting and National Institute for Environmental Studies is responsible for data security and data bank as data center. And medical support center is National Center for Child Health and Development, this is a national hospital and they support medical aspects. 15 regional unit centers located across Japan. Regional unit center coordinate local government and hospitals, clinic and schools, in their area. They collaborate with local stakeholders. These are our study locations. From Hokkaido to Okinawa, there 15 regional unit centers. These two areas, Miyagi and Fukushima unit centers, are affected by the terrible earthquake in 2011 March. Our recruitment plan had changed at these two areas.

And this slide shows our recruitment strategies. In Japan, there are about one millions of newborn children per year. This is a million newborn children in Japan per 1 year, so we recruit 100 participants over 3 years, so 30,000 newborns per year, so 3% of newborn in Japan are recruited to our JECS Study. And our recruitment period is from 2011 to

2014, and one feature of our recruitment is exclude pregnant women who plan to go back to their home country, hometown to deliver called 'SATOGAERI SYUSSAN'.

In Japan, there are some special habit, which is returning to pregnant women's parent's house to give birth. We call it 'SATOGAERI SYUSSAN'. We exclude pregnant women who plan to SATOGAERI because for such pregnant women we cannot get bio-specimen like cord blood at birth. So, bio-specimen at birth is important we think, so we set this exclusion criteria. Exclusion criteria is women who deliver other areas of study zone is excluded.

And this is our recruitment progress. From 2011 January to 2014 March is the recruitment period, so this has dynamically decreased. And this March, we were hit with a terrible earthquake here, so we started before the earthquake but many, many tough problems are here, so not much people could be recruited, but finally, we achieved the goal of our recruitment and we recruited about 103,000 participants at the end of the recruitment.

And at this point, so from here we change recruitment area about Fukushima, because at first plan until here, only five cities in Fukushima is recruitment area but many pregnant women who live in other areas of Fukushima want to join JECS, so we changed our recruitment plan and any pregnant woman who lived in Fukushima area can join our study so we changed our recruitment strategy here, so this is the Fukushima area people were added, so recruitment is increased and increased and finally about 103,000 participants joined in our studies.

And this is concentrate. Yes, concentrate is very good, almost 80% of mothers participate in JECS, 77.7% of mothers concentrate to JECS and more than 90% of fathers, whose wives participated in JECS, agreed to join the study but concentrates vary across study location so, the worst area, 66% and best area 90%, so average is 77% but the concentrate varies across the study locations.

And this slide showed our research components. We conduct informed consent at the first trimester of pregnancy is here, it's quite beginning of the pregnancy, and we conducted questionnaires, survey and collect mother's blood, urine samples, and collect environmental samples in the pregnant period, and at the time of birth, we confirmed health of baby at birth and collect umbilical blood sample, and collect father's blood sample and so on.

After 1 month from birth, breast milk is collected. Every bio-specimen is stored and measure chemical substances. And after 6 months to 13 years, it means the end of 12 years, questionnaires and interview and environmental sample is collected. This is for main study, this is for sub-study, and this two is for sub-study.

All of the data send electrically to the data center. We clean up these data set, and statistical analysis is conducted to identify the factors impacting children's development. We preserve bio-specimen in long term for future study needs. For mothers the blood sample, urine, and breast milk and hair and questionnaire is collected. For fathers, blood samples, urine and questionnaire is collected.

For babies, dried blood spot, cord blood sample, and urine is collected and of course questionnaire is answered by mother or father, but their health status is monitored by questionnaires. So, long-term stock is needed for future study needs and chemical attacks to proteins, metabolomics and screening for disease – this is an important point, so if we find they have some disease and we have to provide the information about these things to their participants.

And this table shows our outcome and exposure measurement schedule. For mothers, T1 means first trimester of pregnancy, T2 means second trimester of pregnancy, so three timing of bio-specimens are collected. Father is only one timing to collect bio-specimen, and for children, at birth; 1 month, 6 months, 12 months, and from 1.5 years to 12 years, every 6 months, and maternal blood, breast milk, urine, hair, paternal blood, cord blood, child blood, child urine, child hair, like this schedule. This pink area is not yet decided officially because our operation is so called 'bicycle operation'.

This slide shows especially about self-administrated questionnaire, I hope you are interested in these things. The self-administrative questionnaire is provided for each 6 months, including lifestyle, mental and physical development. We use about autism, autism screening questionnaire, ASQ, and mental conditions for mothers, and residential environment, and socioeconomic factors like these kind of questionnaire. We read these questionnaires using OCR system and conducted cleaning at data center.

This slide, this figure shows the JECS roadmap and budget. So, our study have 3 years' recruitment period and main study started from 2011 and follow up started from 2014. Biological sample analysis is started from this year. Sub-studies start from next month so now we finished random samples and now called the recruitment, could you join our sub-study, okay or not, now we have called the subject to join the sub study.

Case Cohort has not yet started, but some adjunct study is already started. And pilot study started from 2008, and about 400 children are recruited as pilot study subjects. Data management and data cleaning is almost done and data analysis is started from last year. Now, we are here at this roadmap.

This slide shows our participants' especially mother's mental condition. This is not exposure-disease relationship, it means not causal relationship. This figure shows only cross-sectional and aggregated data. To the question "Do you have any unpleasant feeling

against your child, after child is 1 year old?", most participants say 'never', or 'hardly ever', but some participants, 'almost always' or 'sometimes'. Parcentage of these kind of answer goes higher and higher against mother's age. It suggest that some young mothers need more help to their mental conditions.

Next question is "Do you feel hopeless during this month". Tendency of answer to this question is same as the last question. Most participants say 'never' or 'hardly ever' but some say it's 'usually' or 'almost always' and younger mother feel more hopeless about their children and child care. But, to the question "Do you feel everything tough work during this month", younger mother tend to answer 'never', or 'hardly ever' more than aged mother. We think this is interesting results. To feel unpleasant, to feel hopeless and to feel tough is opposite tendency to mother's age.

This slide shows the future statistical problem. Now, we try to do many simulation study and now we try to modify traditional statistical method to solve JECS's original problem. The first problem is about correlated data. We have 100,000 subjects, but 1% of subject is twins, and few percent of subjects are brothers or sisters, it means both children have same mother and are subjects of JECS. These data must be strongly correlated. So we need to apply the method of 'correlation data analyses' to JECS data.

The second problem is very tough. It is detection limit problem. Some chemicals, we cannot detect the limit of measurement, it is not missing, not zero but under the detection limits. So cannot determine the exact value of these chemicals. It is no problem that few percent of subjects' measurements are under the detection limits, but we expect for some chemicals, more than 50% of the subjects' measurements value is under detection limit. This must be the big statistical problem. Additional to detection limit problem, the value of detection limit is not always the same. The value of detection limit differ between days, laboratories and measurement lots.

And the third statistical problem is complex causal mechanism. JECS data is so complex longitudinal data. This slide shows simple causal diagram and DAG, but actually, real causal pathway must be more complexes. Some exposure cause some outcome, and it cause another exposure or promote other outcomes. Every exposure and outcome are convoluted. Additional to this third problem, we need to divide direct and indirect effect of causality. Direct effect is effect estimate of direct pathway from exposure to outcome, indirect effect is effect estimate of indirect pathway from exposure to outcome through intermediate factors or confounders. We call this kind of estimation of path diagram as "life course epidemiology". We have to use marginal structure model or G-estimation to estimate direct and indirect effect.

This is the final slide. I'm very proud to mention that JECS data will be published as

national database in future. Of course, the publication is after data cleaning and core hypothesis analysis is finished. Thank you for your attention.

Kensuke Okada

Thank you very much. I think this is one of the biggest data that you can achieve in research survey or scientific survey, so do you have any questions or comments?

Questioner

About the big increase in the congenital abnormality from maybe 40 years ago to now, but even compared with the 90s there is a huge increase. Do you expect that this increase is due to a chemical substance? I don't think so actually, from 1990s to 2000s there is not much change, at least I feel. Do you have any guess for this kind of big increase?

Ayano Takeuchi

This is a difficult question, but most exposure scientists say that this is not so small contribution about the chemicals, they say. And of course, I think, the aging of pregnant women's aging higher and higher is the biggest problem I think but the exposure scientists said some chemicals affect big contribution to this increase.

Questioner

That's what you are going to investigate, so it is difficult to answer now but if you have a guess would you tell me the effect size of such kind of environmental chemical factors compared with the age or something like that?

Ayano Takeuchi

Effect size...

Questioner

Of course, I can wait till 10 years or 12 years.

Ayano Takeuchi

In scale risk ratio, I think less than 1.1, I think, I guess.

Kensuke Okada

Any other questions? I think basically for this kind of data, the effect size is very small but still the effect is important and consistent I think. Any other questions or comments? Please.

Questioner

Thank you very much for your nice presentation. My question is, recently I hear surrogation and artificial insemination, so manipulation of the birthing children. So, my question is there are so many deformed children right now, as you mentioned. To avoid that, there are many people who are trying to pretest whether or not their children are deformed or not. So, what is your basic idea about to know before pregnancy, to check the fetus? From what I know, is it a nice or ethical approach to the – is it ethical or good approach, what is your opinion, I want to hear?

Ayano Takeuchi

Alright, I think, you mean if I detect some factor we cannot change beforehand pregnant, how can I treat?

Questioner

We are now in a position to kill children if he or she is mentally, for example, retarded or physically handicapped. In that case, maybe I do not want to have such a child. In that case, I'm in a position to kill before, so it may be a nice example of destroying a human right. So, this is my question.

Ayano Takeuchi

Alright. This is a very big ethical problem I think but I think our study is different from the early detection. Our aim is prevention, so maybe...

Questioner

I hope that the study could be different from that kind of question. I think that kind of question is basically beyond the scientific question, but...

Ayano Takeuchi

Difficult but good question. We think that actionability is important we think so the risk factor, we cannot take any action to detect this kind of risk factor, it's not ethical we think, so this is our main stance. So, actionable risk factor we want to detect.

Kensuke Okada

Alright, any other questions or comments? Alright, so let us give a big hand to Dr. Takeuchi again.