

mmHg, (ii) diastolic blood pressure ≥ 90 mmHg or (iii) being prescribed anti-hypertensive drug(s). Among the 13 321 subjects without self-report-documented hypertension, 3097 had hypertension, i.e. fulfilled criterion (i) and/or (ii) and/or (iii). A subject's family history of diabetes was considered positive if at least one parent or one sibling had diabetes. Body mass index (BMI) was calculated as the weight (kg)/[height (m)]² and used as an index of relative weight. The subjects' weight and height acquired from the questionnaire were validated by the data obtained from the health check-up, which about one-third of the subjects voluntarily underwent [20].

Ascertainment of diabetes mellitus

Whether the subject had a prevalent disease was determined by the questions on the baseline questionnaire: i.e. 'Has a doctor ever told you that you have any of the following diseases?—diabetes (yes/no), hypertension (yes/no)', and so forth. 'Prevalent diabetes' was defined as a reply of 'yes' to the question concerning diabetes. Individuals with diabetes at baseline were excluded from this analysis. Individuals without diabetes at baseline who subsequently answered 'yes' on either or both of the follow-up questionnaires at years 5 and 10 were considered to have developed diabetes. A total of 1183 subjects (703 men and 480 women) reported the development of diabetes during the 10-year study period. We classified all incident cases of diabetes as Type 2, as the age of onset in this middle-aged cohort was 40 years or older.

To document the validity of the self-report, we examined a series of medical records as follows. For practical reasons, three of the 14 administrative districts were chosen to validate the questionnaire information. In these areas, there were 207 participants recorded as having diabetes at year 5. We sent a letter to these participants requesting permission to examine their medical records and 167 replies were received. Of these, 154 participants were confirmed as having diabetes again by self-report. Permission to review their medical records was obtained from 110 of the 154 participants, and the records of 93 participants (54 men and 39 women) of major hospitals were chosen for verification. Two specialists in diabetes (M.N. and Y.T.) reviewed the records, and diabetes was confirmed if any of the following criteria were met: (i) the World Health Organization (1985) criteria [21], (ii) a high casual plasma glucose level (≥ 11 mmol/l), or (iii) use of diabetic medication (insulin or oral hypoglycaemic agent). Thirty subjects (19 men and 11 women) met criterion (i), eight subjects (five men and three women) met criterion (ii), and 38 subjects (20 men and 18 women) met criterion (iii). When we applied the new criteria of the American Diabetes Association (1997) [22], the number of confirmed cases of diabetes did not change, as none of the subjects with a 2-h post-challenge level of < 11 mmol/l had a fasting plasma glucose level in the diabetic range specified by the new criteria alone. In summary, a diagnosis of diabetes was confirmed in a total of 76 of the 93 subjects (82%) who were screened, which we considered reasonable and sufficiently high for a large-scale study. Among the 17 subjects in whom a diagnosis of diabetes was not confirmed, the medical records of 12 subjects were unavailable ($n = 9$) or contained insufficient data to justify a diagnosis of diabetes ($n = 3$). When only subjects for whom complete medical records were available were analysed, the percentage of confirmed diagnosis increased to 94%.

We also conducted a cross-sectional survey to examine whether self-report of diabetes agreed with diagnosis based on health check-up data among Cohort I participants. We collected blood samples from 12 460 subjects (29% of the study cohort) who voluntarily participated in the health check-up examination. Participants were determined to have diabetic hyperglycaemia based on their health check-up data if at least one of the following criteria was met: (i) fasting plasma glucose ≥ 7 mmol/l, (ii) casual plasma glucose ≥ 11 mmol/l, or (iii) HbA_{1c} $\geq 6.1\%$ [23]. In a preliminary analysis, out of 1075 subjects with diabetic hyperglycaemia, 498 reported diabetes; meanwhile, among 11 385 subjects without diabetic hyperglycaemia, 11 169 did not report diabetes. According to these results, the sensitivity and specificity of the questionnaire for diabetic hyperglycaemia was roughly 46% and 98%, respectively. Although these analyses were performed without regard to the self-reported current treatment for diabetes, the number of subjects with a self-report of pharmacological treatment without diabetic hyperglycaemia was very small ($\sim 0.4\%$ of the total number of subjects who were without diabetic hyperglycaemia); therefore, overestimation of specificity by this was likely to be within a negligible range.

Analysis

All analyses were performed separately for men and women. The statistical significance of baseline differences with regard to diabetes status at follow-up in relation to established and suspected risk factors for Type 2 diabetes was assessed using *t*-tests and χ^2 -tests. A *P*-value < 0.05 was considered significant. The cumulative incidence of diabetes over the 10-year period was selected as the outcome, (a) because risk estimates could be calculated directly, and (b) because the lack of precise dates of diabetes onset precluded the use of a person-year approach. The cumulative incidence was defined as the number of new cases of diabetes occurring during the 10-year follow-up period divided by the number of subjects at risk of developing diabetes at baseline. Multiple logistic regression analysis was used to assess the independent contributions of the risk factors to the subsequent risk for Type 2 diabetes and to obtain odds ratios that were adjusted for the other risk factors. Smoking status (four levels), alcohol intake [four levels, ALC_0 consists of non-drinkers (1349 men and 1916 women) and infrequent occasional drinkers who consume alcohol on three or fewer days a month (2449 men and 12 331 women)], physical exercise (active/inactive), family history of diabetes (positive/negative), and prevalent hypertension (positive/negative) were fitted as categorical variables in our logistic model. Because there were no significant interactions between any of the variables and the areas where the subjects resided, the geographical areas were not included as a variable in the final model and all four areas were analysed together. The 95% confidence interval for each odds ratio was calculated. The Mantel extension test was employed to analyse the trend across increasing levels of alcohol consumption. Statistical significance was determined by 95% confidence intervals not including 1.00 for logistic analyses. The statistical analyses were performed using SAS software (version 8.2; SAS Institute Inc., Cary, NC, USA).

To examine the possible existence of a significant interaction between alcohol consumption and BMI with regard to the risk

of diabetes, we conducted a stratified analysis for BMI with cut-off levels set at 22 and 25 kg/m²; these values represent the ideal BMI and the lower BMI limit of obesity, respectively, for Japanese people as defined by the Japan Society for the Study of Obesity [24]. The former value was determined by the BMI associated with the lowest level of morbidity among middle-aged Japanese [25].

Results

Incident Type 2 diabetes mellitus (Table 1)

During the 10-year follow-up, we documented 703 incident cases (5.4%) of diabetes among men and 480 cases (3.0%) among women. There was male predominance in the incidence of diabetes.

Risk factors for diabetes at baseline (Table 2)

Subjects of both genders who converted to a diabetes-positive status were significantly older and had a higher BMI than those who remained non-diabetic. In addition, higher percentages of subjects were positive for smoking, family history of diabetes and past history of hypertension among those who became diabetic during the follow-up period than among those who remained non-diabetic. The percentage of men with moderate (ethanol intake: > 23 g/day and ≤ 46 g/day) or high (ethanol intake > 46 g/day) alcohol consumption was also higher among subjects who became diabetic during the follow-up compared with those who remained non-diabetic. There was an increasing trend for developing diabetes during the follow-up period according to alcohol consumption, and this positive trend was significant (*P* for trend = 0.007) by the Mantel extension test.

Table 1 Ten-year incidence of Type 2 diabetes mellitus in the JPHC Cohort according to gender

Age (years)	Men		Women	
40–49	309/6404	(4.8)	191/7698	(2.5)
	80/1471	(5.4)	52/1951	(2.7)
	80/1835	(4.4)	47/2230	(2.1)
	90/1900	(4.7)	47/2069	(2.3)
	59/1198	(4.9)	45/1448	(3.1)
50–59	394/6509	(6.1)	289/8282	(3.5)
	92/1386	(6.6)	71/1939	(3.7)
	98/1889	(5.2)	91/2650	(3.4)
	118/1955	(6.0)	81/2251	(3.6)
	86/1279	(6.7)	46/1442	(3.2)
Total	703/12 913	(5.4)	480/15 980	(3.0)

Data are incidence/total number and the per cent (in parentheses). Below the total number and per cent, incidence of diabetes and the per cent of each subcohort are shown. The data are shown for (top to bottom) the Ninohe PHC area of Iwate Prefecture, the Yokote PHC area of Akita Prefecture, the Saku PHC area of Nagano Prefecture, and the Ishikawa PHC area of Okinawa Prefecture.

BMI, family history of diabetes, smoking and risk of diabetes (Table 3)

Multiple logistic regression analysis was performed to determine which of the baseline characteristics that had been previously identified as risk factors in some of the earlier studies were independent predictors of diabetes in the present cohort. Age, BMI, a positive family history of diabetes and a past history of hypertension were strong predictors for the development of diabetes in both genders. Smoking status was also strongly associated with the development of future diabetes among former smokers and those smoking 20 cigarettes or more a day in both genders.

Alcohol consumption and risk of diabetes

Among men, daily alcohol consumption of 23 g of ethanol or more was significantly related to the future development of diabetes when compared with the group of non-drinkers and infrequent occasional drinkers; a positive trend across the increasing levels of alcohol consumption was also significant (*P* for trend = 0.019) according to the Mantel extension test (Table 3).

To determine whether the BMI modified the association between daily alcohol consumption and the risk of Type 2 diabetes, we stratified the subjects according to the BMI (see Table 4). Among lean men (BMI ≤ 22 kg/m²), a significant and strong positive association with moderate to high alcohol consumption was observed and the positive trend across the increasing levels of alcohol consumption was also significant (*P* for trend < 0.001). The risk for heavy alcohol drinkers was 2.89 (95% CI, 1.63–5.11) times higher than that of non-drinkers and infrequent occasional drinkers. By contrast, among men with a BMI > 22 kg/m², only a small, non-significant increase was observed among alcohol consumers (Table 4). In addition, when we analysed non-drinkers and infrequent occasional drinkers separately in the analysis shown in Table 3 (i.e. without subdividing the subjects according to their BMI), the odds between these two groups were almost equal [odds ratio for the former to the latter: 1.01 (95% CI, 0.74–1.38)], with the significantly increased odds ratios for high (> 46 g/day) alcohol consumption compared with non- or occasional infrequent drinkers persisted in the lower (≤ 22 kg/m²) BMI group, even in this stratified analysis (data not shown).

No significant association between alcohol intake and the future development of diabetes was observed among women.

Discussion

This study is the largest community-based prospective study in Japan with a 10-year follow-up period to quantify the risk factors for Type 2 diabetes. We identified established risk factors, such as age, BMI and family history of diabetes, as independent determinants of Type 2 diabetes in both men and women, consistent with the results of studies in western populations [26–34].

Table 2 Baseline characteristics and development of Type 2 diabetes mellitus in middle-aged Japanese men and women

Characteristics	Men			Women		
	Remained non-diabetic	Developed diabetes	P	Remained non-diabetic	Developed diabetes	P
Age (years)	49.4 49.0, 49.5 49.5, 49.4	50.1 49.8, 50.0 50.1, 50.5	0.002	49.6 49.3, 49.9 49.7, 49.2	50.8 50.3, 51.5 51.3, 49.9	< 0.001
BMI (kg/m ²)	23.4 23.5, 23.2 23.1, 24.4	25.0 25.2, 24.8 24.4, 25.9	< 0.001	23.5 23.6, 23.2 23.1, 24.2	25.6 25.6, 25.1 25.2, 26.7	< 0.001
Smoking status (%)			0.012			< 0.001
Never smokers	25.2 27.8, 23.2 21.3, 31.5	21.3 25.0, 22.5 15.4, 24.1		94.6 95.6, 96.0 92.7, 93.6	90.8 93.5, 93.5 92.2, 81.3	
Current smokers:						
1–19 cigarettes/day	15.2 16.7, 15.9 15.6, 11.9	13.2 15.7, 10.7 14.4, 11.7		3.1 2.8, 2.3 4.5, 2.9	2.9 0.8, 2.9 3.9, 4.4	
≥ 20 cigarettes/day	36.7 38.6, 38.3 38.8, 29.2	38.7 41.3, 39.3 39.9, 33.1		1.0 0.7, 0.8 1.1, 1.6	2.5 1.6, 1.5 0.8, 7.7	
Past smokers	22.8 17.0, 22.6 24.3, 27.4	26.7 18.0, 27.5 30.3, 31.0		1.3 1.0, 0.9 1.7, 1.9	3.8 4.1, 2.2 3.1, 6.6	
Alcohol intake* (%)			0.046			0.824
ALC_0	31.4 34.8, 21.8 28.0, 47.2	27.9 29.7, 14.0 22.6, 50.3		89.8 91.9, 87.2 87.0, 95.7	90.8 93.5, 88.4 87.5, 95.6	
ALC_1	25.9 23.4, 24.6 28.7, 26.4	24.0 25.0, 20.8 25.0, 25.5		3.0 2.2, 3.7 4.4, 0.9	3.1 1.6, 3.6 5.5, 1.1	
ALC_2	22.4 20.5, 27.6 25.2, 12.4	24.8 21.5, 34.8 27.4, 12.4		4.1 3.3, 5.1 5.1, 1.8	3.3 2.4, 4.4 3.9, 2.2	
ALC_3	20.3 21.3, 26.1 18.1, 14.0	23.3 23.8, 30.3 25.0, 11.7		3.1 2.6, 3.9 3.6, 1.6	2.7 2.4, 3.6 3.1, 1.1	
Leisure-time physical activity at least once a week (%)	17.2 11.7, 15.4 19.4, 22.6	16.4 13.6, 17.4 13.0, 22.8	0.588	14.2 8.0, 11.8 20.4, 17.3	15.2 11.4, 12.3 19.5, 18.7	0.528
Family history of diabetes (%)	8.2 10.4, 8.6 8.2, 5.0	15.1 18.6, 12.4 17.3, 11.0	< 0.001	8.1 8.7, 8.4 8.9, 5.9	18.8 25.2, 16.7 21.1, 9.9	< 0.001
History of hypertension (%)	15.0 15.7, 18.1 13.6, 11.6	22.5 26.7, 26.4 20.2, 15.9	< 0.001	13.9 15.1, 15.3 13.1, 11.5	29.0 34.2, 31.2 23.4, 26.4	< 0.001

Data are means (age and BMI) or percentages (all others).

*Alcohol intake (g/day of ethanol): men, ALC_1: 0 < ethanol ≤ 23.0, ALC_2: 23.0 < ethanol ≤ 46.0, ALC_3: ethanol > 46.0; women, ALC_1:

0 < ethanol ≤ 4.9, ALC_2: 4.9 < ethanol ≤ 11.5, ALC_3: ethanol > 11.5. ALC_0: non-drinkers and infrequent occasional drinkers who consume alcohol on three or fewer days a month.

The total data and data for each subcohort are shown. Data are shown for (left to right, top to bottom) the Ninohe PHC area of Iwate Prefecture, the Yokote PHC area of Akita Prefecture, the Saku PHC area of Nagano Prefecture, and the Ishikawa PHC area of Okinawa Prefecture.

The analysis revealed a significant positive association between moderate to high alcohol intake and future diabetes in lean men (BMI ≤ 22 kg/m²) and a similar but non-significant correlation in obese men (BMI > 22 kg/m²). This contrasts with the results for men in most previous studies in the United States and Europe conducted using a prospective design, which reported

an inverse correlation between alcohol intake and Type 2 diabetes [33,35,36] or suggested no significant association with diabetes [37–39]. A few, however, showed an excess diabetes incidence only in heavy drinkers [40,41].

The results of the Osaka Health Survey of Japanese male employees [15] showed moderate alcohol consumption (21.1–

	Men (<i>n</i> = 12 913)		Women (<i>n</i> = 15 980)	
	Odds ratio (95% CI)		Odds ratio (95% CI)	
Age (1-year increase)	1.02	(1.01–1.04)	1.02	(1.01–1.04)
BMI (1 kg/m ² -increase)	1.17	(1.14–1.20)	1.17	(1.14–1.21)
Smoking status				
Never smokers	1.00 (referent)		1.00 (referent)	
Current smokers:				
1–19 cigarettes/day	1.14	(0.87–1.50)	1.07	(0.62–1.86)
≥ 20 cigarettes/day	1.37	(1.11–1.69)	2.94	(1.57–5.50)
Past smokers	1.35	(1.08–1.69)	2.77	(1.67–4.61)
Alcohol intake*				
ALC_0	1.00 (referent)		1.00 (referent)	
ALC_1	1.08	(0.87–1.34)	1.15	(0.68–1.95)
ALC_2	1.26	(1.02–1.56)	0.81	(0.48–1.35)
ALC_3	1.25	(1.00–1.56)	0.78	(0.44–1.40)
Family history (yes/no)	2.00	(1.60–2.49)	2.69	(2.12–3.43)
Leisure time physical activity (active/inactive)	0.90	(0.73–1.12)	1.06	(0.82–1.37)
Hypertension (yes/no)	1.34	(1.10–1.62)	1.79	(1.44–2.22)

* Alcohol intake (g/day of ethanol): men, ALC_1: 0 < ethanol ≤ 23.0, ALC_2: 23.0 < ethanol ≤ 46.0, ALC_3: ethanol > 46.0; women, ALC_1: 0 < ethanol ≤ 4.9, ALC_2: 4.9 < ethanol ≤ 11.5, ALC_3: ethanol > 11.5. ALC_0: non-drinkers and infrequent occasional drinkers who consume alcohol on three or fewer days a month. 95% CI, 95% confidence interval.

Table 3 Multivariate logistic regression analysis of the 10-year incidence of Type 2 diabetes mellitus in middle-aged Japanese according to gender

Table 4 Multivariate logistic regression analysis of the 10-year incidence of Type 2 diabetes mellitus in middle-aged Japanese males according to BMI

	BMI ≤ 22 kg/m ² (<i>n</i> = 3845)		25 kg/m ² ≥ BMI > 22 kg/m ² (<i>n</i> = 5671)		BMI ≥ 25 kg/m ² (<i>n</i> = 3397)	
	Odds ratio (95% CI)		Odds ratio (95% CI)		Odds ratio (95% CI)	
Alcohol intake*						
ALC_0	1.00 (referent)		1.00 (referent)		1.00 (referent)	
ALC_1	1.05	(0.55–2.01)	1.12	(0.80–1.56)	1.08	(0.79–1.48)
ALC_2	1.91	(1.05–3.46)	1.16	(0.83–1.61)	1.24	(0.89–1.71)
ALC_3	2.89	(1.63–5.11)	1.17	(0.83–1.66)	1.03	(0.73–1.44)

* Alcohol intake (g/day of ethanol): ALC_1: 0 < ethanol ≤ 23.0, ALC_2: 23.0 < ethanol ≤ 46.0, ALC_3: ethanol > 46.0. 95% CI, 95% confidence interval. Adjusted for age, BMI, cigarette smoking, exercise, family history of diabetes and prevalent hypertension.

50.0 ml/day) to be associated with reduced risk of Type 2 diabetes among men with BMI ≥ 22.1 kg/m², while heavy alcohol consumption was associated with an increased risk among lean men (BMI ≤ 22.0 kg/m²). All the subjects of the Osaka Health Survey were employees of the same company, while the subjects of our study were composed of those living in several areas of Japan. Therefore, the study population in our analysis may be more representative of the general Japanese. However, the heterogeneity of socio-economic status in our cohort could not completely exclude potential confounding on, for example, alcohol consumption.

Recently, three more reports have been published that deal with the relationship between alcohol consumption and the risk of Type 2 diabetes among Japanese [42–44]. One of these reported a significant protective effect of a low level of alcohol consumption (23.0–45.9 g ethanol/day) against development of Type 2 diabetes, i.e. a possible U-shaped association among

male employees during 7 years of follow-up [42]. Another study demonstrated a significantly positive association in lean (BMI ≤ 22.0 kg/m²) but no significant association in obese (BMI ≥ 25.0 kg/m²) subjects; and a significant negative association in those who had an intermediate BMI (22.1–24.9 kg/m²) between current alcohol consumption and the incidence of Type 2 diabetes, following male (72%) and female (28%) employees for a mean of 5.7 years [43]. In addition, the Hisayama Study identified alcohol consumption as an independent risk factor for diabetes among males [44]. Summing up the literature and our data, alcohol consumption exceeding 46 g/day is concluded to have an unfavourable effect, prompting Type 2 diabetes development, especially among lean (BMI ≤ 22.0 kg/m²) Japanese men. The apparent lack of an association in women in the present study may be due to the small number of alcohol drinkers among the women surveyed (Table 2).

No significant correlation was found between leisure-time physical activity and diabetes development, a finding somewhat different from other prospectively designed studies, most of which showed a significant association between physical inactivity and Type 2 diabetes [30,37,45–47]. In this context, it may have been a limitation in regard to assessing the association that we categorized subjects only according to frequency of leisure-time exercise in the present study.

Our study has some limitations. First, only self-reported information was available regarding the subject's diabetes status. Although the validation examination showed that self-reported diabetes reflected the true situation fairly well (more than 80%) in this general population, the number of those with undiagnosed diabetes who were in the non-diabetic category according to self-report should be estimated. For this purpose, we compared the number of self-reported diabetic subjects with the number of those who were diagnosed on the basis of plasma glucose and HbA_{1c} levels [23] in a group of approximately 14 000 health check-up examinees involved in the JPHC Cohort I. The results showed that about half of all prevalent cases (self-reported and blood-sample diagnosed combined) were undiagnosed until the health check-up. Thus, the self-report-defined non-diabetic category at follow-up is likely to have contained a substantial undiagnosed population, possibly similar in number to the diabetic category defined by self-report. This implies that the odds ratios observed in our analyses may have underestimated the effect of risk factors on the total incidence of diabetes.

Second, there may be follow-up bias between the diabetic and non-diabetic categories resulting from a presumed excess mortality of the diabetic patients during the follow-up and a possible altered response rate of the patients. We divided the total number of subjects analysed (i.e. those who replied to the baseline questionnaire) into two groups: those who responded to both follow-up questionnaires (74.5% of the initial respondents) and those who did not, and compared the two groups. There were no significant differences in representative parameters, such as BMI and lifestyle characteristics, in either men or women, and therefore estimated incidence of diabetes calculated on the basis of these parameters. This suggests that our results were not seriously affected by follow-up bias.

Third, data for alcohol consumption were obtained from self-reports. Therefore, there might be under-reporting of true alcohol consumption. In this regard, it should be commented that the levels of γ -glutamyltranspeptidase of the subjects of the group where risk for diabetes starts to increase are roughly estimated to have been ranged between -30 and -80 IU/l as a whole [19], which correspond to normal to moderately high levels of γ -glutamyltranspeptidase of Japanese male population. Finally, previously diagnosed medical conditions were self-reported by participants, so this study may not exclude subjects with asymptomatic chronic alcoholic liver disease at baseline.

The average annual incidence of Type 2 diabetes by self-report in the current study was calculated as 0.63% among men and 0.34% among women, incidences that, including

surmised undiagnosed cases, lie in the lower middle range of the reported crude incidence rate in the Japanese general population (0.2–4.0% per year for both men and women) [48], although age range and diagnostic criteria were different from those of our study. The male predominance of the observed incidence in the present study is another interesting point. Similar results were also obtained in the Japanese Governmental investigations of diabetes conducted in 1997 [49] and in 2002 [50], which were based on 5883 and 5792 subjects from among the participants in the National Nutritional Survey of the year, respectively.

In conclusion, most variables predicting future diabetes in western populations were also found to be important predictors of the disease in our current analyses. However, greater emphasis should be placed on alcohol consumption, as it might have more of an adverse than a beneficial effect on development of diabetes, in comparison with western populations. This may be due to the difference in distribution of polymorphic ethanol-metabolizing enzymes between Japanese and western populations [51,52].

Acknowledgements

This study was supported by a grant-in-aid for Cancer Research and for the Second Term Comprehensive Ten-Year Strategy for Cancer Control, and Health Sciences Research grants (Medical Frontier Strategy Research H13-008, Clinical Research for Evidence-based Medicine H14-008) from the Ministry of Health, Labour and Welfare of Japan.

Study group members

The investigators and participating institutions in the JPHC Study Cohort I Group, a part of JPHC Study Group (principal investigator: S. Tsugane) were: S. Tsugane, S. Sasaki, Epidemiology and Biostatistics Division, National Cancer Center Research Institute East, Kashiwa; J. Ogata, S. Baba, National Center for Circulatory Diseases, Suita; K. Miyakawa, F. Saito, A. Koizumi, Iwate Prefectural Ninohe Public Health Center, Ninohe; Y. Miyajima, N. Suzuki, S. Nagasawa, Akita Prefectural Yokote Public Health Center, Yokote; H. Sanada, Y. Hatayama, F. Kobayashi, H. Uchino, Y. Shirai, T. Kondo, Nagano Prefectural Saku Public Health Center, Saku; Y. Kishimoto, E. Takara, M. Kinjo, T. Fukuyama, Okinawa Prefectural Ishikawa Public Health Center, Ishikawa; S. Matsushima, S. Natsukawa, Saku General Hospital, Usuda; S. Watanabe, M. Akabane, Tokyo University of Agriculture, Tokyo; M. Konishi, Ehime University, Matsuyama; S. Tominaga, Aichi Cancer Center Research Institute, Nagoya; M. Iida, S. Sato, Center for Adult Diseases, Osaka; the late M. Yamaguchi, Y. Matsumura, National Institute of Health and Nutrition, Tokyo; Y. Tsubono, Tohoku University, Sendai; H. Iso, Tsukuba University, Tsukuba; H. Sugimura, Hamamatsu University, Hamamatsu; and M. Kabuto, National Institute for Environmental Studies, Tsukuba.

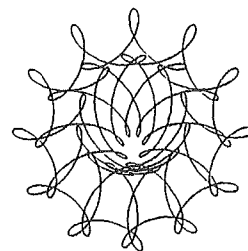
References

- 1 Elbein SC. Perspective: the search for genes for type 2 diabetes in the post-genome era. *Endocrinology* 2002; 143: 2012–2018.
- 2 Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002; 346: 393–403.
- 3 Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* 2001; 344: 1343–1350.
- 4 Fujimoto WY, Leonetti DL, Kinyoun JL, Newell-Morris L, Shuman WP, Stolov WC et al. Prevalence of diabetes mellitus and impaired glucose tolerance among second-generation Japanese-American men. *Diabetes* 1987; 36: 721–729.
- 5 Hara H, Egusa G, Yamakido M, Kawate R. The high prevalence of diabetes mellitus and hyperinsulinemia among the Japanese-Americans living in Hawaii and Los Angeles. *Diabetes Res Clin Pract* 1994; 24: S37–42.
- 6 Tsugane S, Gotlieb SL, Laurenti R, Souza JM, Watanabe S. Mortality and cause of death among first-generation Japanese in Sao Paulo, Brazil. *Int J Epidemiol* 1989; 18: 647–651.
- 7 Meguro M, Meguro K, Caramelli P, Ishizaki J, Ambo H, Chubaci RY et al. Elderly Japanese emigrants to Brazil before World War II. I. Clinical profiles based on specific historical background. *Int J Geriatr Psychiatry* 2001; 16: 768–774.
- 8 King H, Aubert RE, Herman WH. Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections. *Diabetes Care* 1998; 21: 1414–1431.
- 9 Park Y, Lee H, Koh CS, Min H, Yoo K, Kim Y et al. Prevalence of diabetes and IGT in Yonchon County, South Korea. *Diabetes Care* 1995; 18: 545–548.
- 10 Pan XR, Yang WY, Li GW, Liu J. Prevalence of diabetes and its risk factors in China, 1994. National Diabetes Prevention and Control Cooperative Group. *Diabetes Care* 1997; 20: 1664–1669.
- 11 Kawakami N, Takatsuka N, Shimizu H, Ishibashi H. Effects of smoking on the incidence of non-insulin-dependent diabetes mellitus. Replication and extension in a Japanese cohort of male employees. *Am J Epidemiol* 1997; 145: 103–109.
- 12 Nakanishi N, Nakamura K, Matsuo Y, Suzuki K, Tataru K. Cigarette smoking and risk for impaired fasting glucose and type 2 diabetes in middle-aged Japanese men. *Ann Intern Med* 2000; 133: 183–191.
- 13 Okada K, Hayashi T, Tsumura K, Suematsu C, Endo G, Fujii S. Leisure-time physical activity at weekends and the risk of Type 2 diabetes mellitus in Japanese men: the Osaka Health Survey. *Diabet Med* 2000; 17: 53–58.
- 14 Uchimoto S, Tsumura K, Hayashi T, Suematsu C, Endo G, Fujii S et al. Impact of cigarette smoking on the incidence of Type 2 diabetes mellitus in middle-aged Japanese men: the Osaka Health Survey. *Diabet Med* 1999; 16: 951–955.
- 15 Tsumura K, Hayashi T, Suematsu C, Endo G, Fujii S, Okada K. Daily alcohol consumption and the risk of type 2 diabetes in Japanese men: the Osaka Health Survey. *Diabetes Care* 1999; 22: 1432–1437.
- 16 Hayashi T, Tsumura K, Suematsu C, Endo G, Fujii S, Okada K. High normal blood pressure, hypertension, and the risk of type 2 diabetes in Japanese men. The Osaka Health Survey. *Diabetes Care* 1999; 22: 1683–1687.
- 17 Shin CS, Lee HK, Koh CS, Kim YI, Shin YS, Yoo KY et al. Risk factors for the development of NIDDM in Yonchon County, Korea. *Diabetes Care* 1997; 20: 1842–1846.
- 18 Tsugane S, Sobue T. Baseline survey of JPHC study—design and participation rate. Japan Public Health Center-based Prospective Study on Cancer and Cardiovascular Diseases. *J Epidemiol* 2001; 11: S24–29.
- 19 Tsugane S, Fahey MT, Sasaki S, Baba S. Alcohol consumption and all-cause and cancer mortality among middle-aged Japanese men: seven-year follow-up of the JPHC study Cohort I. Japan Public Health Center. *Am J Epidemiol* 1999; 150: 1201–1207.
- 20 Tsugane S, Sasaki S, Tsubono Y. Under- and overweight impact on mortality among middle-aged Japanese men and women: a 10-year follow-up of JPHC study cohort I. *Int J Obes Relat Metab Disord* 2002; 26: 529–537.
- 21 World Health Organization. *Diabetes Mellitus*. Report of a WHO Study Group. Technical Report Series No. 727. Geneva: World Health Organization, 1985.
- 22 The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the Expert Committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 1997; 20: 1183–1197.
- 23 Takahashi Y, Noda M, Tsugane S, Kuzuya T, Iro C, Kadowaki T. Prevalence of diabetes estimated by plasma glucose criteria combined with standardized measurement of HbA_{1c} among health checkup participants on Miyako Island, Japan. *Diabetes Care* 2000; 23: 1092–1096.
- 24 The Examination Committee of Criteria for 'Obesity Disease' in Japan, Japan Society for the Study of Obesity. New criteria for 'obesity disease' in Japan. *Circ J* 2002; 66: 987–92.
- 25 Tokunaga K, Matsuzawa Y, Kotani K, Keno Y, Kobatake T, Fujioka S et al. Ideal body weight estimated from the body mass index with the lowest morbidity. *Int J Obes* 1991; 15: 1–5.
- 26 Wilson PW, Anderson KM, Kannel WB. Epidemiology of diabetes mellitus in the elderly. The Framingham Study. *Am J Med* 1986; 80: 3–9.
- 27 Njolstad I, Arnesen E, Lund-Larsen PG. Sex differences in risk factors for clinical diabetes mellitus in a general population: a 12-year follow-up of the Finnmark Study. *Am J Epidemiol* 1998; 147: 49–58.
- 28 Meisinger C, Thorand B, Schneider A, Stieber J, Doring A, Lowel H. Sex differences in risk factors for incident type 2 diabetes mellitus: the MONICA Augsburg cohort study. *Arch Intern Med* 2002; 162: 82–89.
- 29 Burchfiel CM, Curb JD, Rodriguez BL, Yano K, Hwang LJ, Fong KO et al. Incidence and predictors of diabetes in Japanese-American men. The Honolulu Heart Program. *Ann Epidemiol* 1995; 5: 33–43.
- 30 Helmrich SP, Ragland DR, Leung RW, Paffenbarger RS Jr. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *N Engl J Med* 1991; 325: 147–152.
- 31 Wei M, Gibbons LW, Mitchell TL, Kampert JB, Lee CD, Blair SN. The association between cardiorespiratory fitness and impaired fasting glucose and type 2 diabetes mellitus in men. *Ann Intern Med* 1999; 130: 89–96.
- 32 Feskens EJ, Kromhout D. Cardiovascular risk factors and the 2.5-year incidence of diabetes mellitus in middle-aged men. The Zutphen Study. *Am J Epidemiol* 1989; 130: 1101–1108.
- 33 Rimm EB, Chan J, Stampfer MJ, Colditz GA, Willett WC. Prospective study of cigarette smoking, alcohol use, and the risk of diabetes in men. *BMJ* 1995; 310: 555–559.
- 34 Rimm EB, Manson JE, Stampfer MJ, Colditz GA, Willett WC, Rosner B et al. Cigarette smoking and the risk of diabetes in women. *Am J Public Health* 1993; 83: 211–214.
- 35 Ajani UA, Hennekens CH, Spelsberg A, Manson JE. Alcohol consumption and risk of type 2 diabetes mellitus among US male physicians. *Arch Intern Med* 2000; 160: 1025–1030.
- 36 Stampfer MJ, Colditz GA, Willett WC, Manson JE, Arky RA, Hennekens CH et al. A prospective study of moderate alcohol drinking and risk of diabetes in women. *Am J Epidemiol* 1988; 128: 549–558.
- 37 James SA, Jamjoum L, Raghunathan TE, Strogatz DS, Furth ED, Khazanie PG. Physical activity and NIDDM in African-Americans. The Pitt County Study. *Diabetes Care* 1998; 21: 555–562.
- 38 Ohlson LO, Larsson B, Bjorntorp P, Eriksson H, Svardsudd K, Welin L et al. Risk factors for type 2 (non-insulin-dependent) diabetes mellitus. Thirteen and one-half years of follow-up of the participants in a study of Swedish men born in 1913. *Diabetologia* 1988; 31: 798–805.

- 39 Perry IJ, Wannamethee SG, Walker MK, Thomson AG, Whincup PH, Shaper AG. Prospective study of risk factors for development of non-insulin dependent diabetes in middle aged British men. *Br Med J* 1995; 310: 560–564.
- 40 Holbrook TL, Barrett-Connor E, Wingard DL. A prospective population-based study of alcohol use and non-insulin-dependent diabetes mellitus. *Am J Epidemiol* 1990; 132: 902–909.
- 41 Kao WH, Puddey IB, Boland LL, Watson RL, Brancati FL. Alcohol consumption and the risk of type 2 diabetes mellitus: atherosclerosis risk in communities study. *Am J Epidemiol* 2001; 154: 748–757.
- 42 Nakanishi N, Suzuki K, Tatara K. Alcohol consumption and risk for development of impaired fasting glucose or type 2 diabetes in middle-aged Japanese men. *Diabetes Care* 2003; 26: 48–54.
- 43 Watanabe M, Barzi F, Neal B, Ueshima H, Miyoshi Y, Okayama A *et al*. Alcohol consumption and the risk of diabetes by body mass index levels in a cohort of 5636 Japanese. *Diabetes Res Clin Pract* 2002; 57: 191–197.
- 44 Kiyohara Y, Shinohara A, Kato I, Shiota T, Kubo M, Tanizaki Y *et al*. Dietary factors and development of impaired glucose tolerance and diabetes in a general Japanese population: the Hisayama study. *J Epidemiol* 2003; 13: 251–258.
- 45 Manson JE, Nathan DM, Krolewski AS, Stampfer MJ, Willett WC, Hennekens CH. A prospective study of exercise and incidence of diabetes among US male physicians. *JAMA* 1992; 268: 63–67.
- 46 Manson JE, Rimm EB, Stampfer MJ, Colditz GA, Willett WC, Krolewski AS *et al*. Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet* 1991; 338: 774–778.
- 47 Folsom AR, Kushi LH, Hong CP. Physical activity and incident diabetes mellitus in postmenopausal women. *Am J Public Health* 2000; 90: 134–138.
- 48 Akazawa Y. Prevalence and incidence of diabetes mellitus by WHO criteria. *Diabetes Res Clin Pract* 1994; 24: S23–27.
- 49 Editorial Board. Report of the investigation for diabetes in Japan. *J Jp Diabetes Soc* 1998; 41: 325–331. Available from: <http://www1.mhlw.go.jp/toukei/tounyou/>.
- 50 Ministry of Health, Labour and Welfare. Report of the investigation for diabetes in Japan, 2002. Available from: <http://www.mhlw.go.jp/shingi/2003/08/s0806-4.html>.
- 51 Shibuya A, Yoshida A. Frequency of the atypical aldehyde dehydrogenase-2 gene (ALDH2(2)) in Japanese and Caucasians. *Am J Hum Genet* 1988; 43: 741–743.
- 52 Sun F, Tsuritani I, Yamada Y. Contribution of genetic polymorphisms in ethanol-metabolizing enzymes to problem drinking behavior in middle-aged Japanese men. *Behav Genet* 2002; 32: 229–236.

臨床統計学

臨床試験を中心として



大橋 靖雄

臨床統計学 (Clinical Biostatistics) は、健康・医療に関わる応用統計学 Biostatistics (生物統計学と訳される) の重要な 1 分野である。わが国でも生物統計学の重要性がようやく認識され、実務家・研究者の需要が立ち上がるとともに教育システムも生まれつつある。大学では 1992 年に東京大学に初めての講座が誕生し、1999 年の北里大学薬学系研究科、2000 年の京都大学医学系研究科と講座開設が続いた。2002 年には東京理科大学大学院 (経営工学) に社会人対象修士コースが生まれ、2004 年には久留米大学、厚生労働省保健科学院にも教育コースが開設予定である。現時点で生物統計家の参画が最も必要とされているのは、臨床研究とくに臨床試験の計画と解析である。本稿では、歴史的な展開も含め、詳細な数理というよりは、このような専門家が必要となってきた背景の記述と主要な概念について述べてみたい。

1. 臨床医学と EBM, 臨床試験

臨床医学の目的は、患者の疾患を正確に診断し適切に治療を行うことにある。治療目的は、可能な場合には治癒を目指し、治癒が不可能な場合には患者の QOL を可能な限り向上させるか維持することである。しかし、医療提供者側の知識・技術の不完全さ、主に患者・疾患の多様さに多く由来する治療効果の不確実さから、診断結果や治療結果には永遠に除去不可能な曖昧さが伴う。したがって、診断・治療によってもたらされるベネフィットと被るリスクは、ともに「可能性」として確率変数的性格を帯びる。これらに対する患者の重み付けも患者個々の価値判断を反映して異なるはずのものであり、診断・治療法の選択は、リスク・ベネフィット両者のバランスと資源の制約の中で、本来は十分な

情報提示と理解、そして自発的意志を前提としたインフォームドコンセントによるべきである (とされる)。しかし、医療提供者と患者の有する情報の不均衡、意志・患者双方の意識の問題もあり、これまでの治療上の意志決定は、医師主導のパターナリズムの中で、曖昧な状況下で行われてきたとあってよい。近年の情報公開あるいは患者の権利主張の流れは、このような意思決定プロセスに大きな変革を与えようとしている。厚生労働省は、ここ数年、疾患毎に標準治療をまとめたガイドライン策定を各関連学会に依頼し、数多くの疾患ガイドラインやその案が発表されてきた。これは、国民皆保険のもと、治療結果をマスとして評価するシステムが存在しないまま、医師の自由裁量のもと出来高払いで治療が行われてきた反省と、また上記の情報開示の要求に応えるためでもあった。そして、このガイドライン策定過程で (医療関係者には周知であったが) 改めて次の事態が浮き彫りになった。

「わが国には客観的証拠 (evidence) が無い！」

臨床医学系学会では、最近 5 年ほど Evidence Based Medicine (略して EBM) という言葉が大流行である。EBM とは、目の前の患者の問題点を一定の手順で定型化し、主に文献検索と抽出された文献の批判的吟味により過去の「証拠・根拠」を点検し、そこから有効な情報を引き出し、目の前の患者に対して、その特異性を充分考慮しつつ、ときにはこれまでの経験によって修正も加え実践することである。科学と経験の融合である。ここで、evidence を提供するのが患者を対象とした臨床研究成果である。

臨床研究は、患者の診断・治療経過をまとめた症例報告、多数の患者の治療実態下での観察研究、そして臨床試験に大まかに分けられる。薬・手術・放射線などの治

療（予防）手段を研究者側が前向きに制御して行われる実験的研究が臨床試験である。これらの手段をまとめて介入（intervention）と呼ぶ。複数の臨床試験を統計的に併合する「メタアナリシス（meta-analysis）」も、もとの臨床試験の質が高ければ、高い質のevidenceを提供し、ガイドライン策定に大いに利用されている。

臨床試験はヒトを対象とした実験ではあるが、ナチスあるいはわが国の石井部隊（731部隊）が戦時中に行ってきた人体実験とは、倫理性の確保という点で大いに異なる。このための国際合意の実施基準が後述のGCP（Good Clinical Practice）である。

さて、治療法の開発とくに新薬開発に関して一般国民が抱いている最大の誤解は、画期的な新薬が動物実験や試験管内の基礎研究から生れる、という「幻想」である。すべての新薬の認可あるいは既存薬剤の適応拡大には、適切に計画・管理実施され、通常は適切な比較対照を有する（これをadequately well-controlledと呼ぶ）臨床試験が必須である（この事情は医療機器についても同様であるが、簡単のため以下は「薬」「製薬会社」で代表させる）。

GCPに従った臨床試験は、図1に示すように入念な準備と複雑な手順、多くの人々の参加・協力により実施される。事前にプロトコルと呼ばれる研究計画書が策定される。これは研究の意義（科学的側面と倫理的側面）と実施マニュアルの両面を持つ基本文書であり、この中で試験の目的・対象者・介入手段・統計解析を含む評価方法が規定される。このプロトコルに従って試験が進行するために、研究スポンサー（後述する「治験」では製薬会社）あるいは実施施設（病院）は手順書（Standard Operational Procedure; SOP）を定め、人的配置を行う。製薬会社や臨床試験の請負会社

- ◆ 研究計画書（プロトコル）
- ◆ 実施システムとくに人的配置と手順書（SOP）
- ◆ 統計家のインプットと解析計画書
- ◆ CRF（調査票）とその標準化
- ◆ データマネジメントのシステム
- ◆ モニタリングと監査体制
- ◆ 評価基準（有効性と安全性）
- ◆ 品質保証

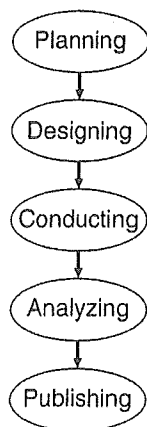


図1 臨床試験に必要な準備。

（Contract Research Organization; CRO）の社員が施設を訪問して実施状況と報告の正確さを点検する活動がモニタリングであり、さらに第三者の監査等を通じて、得られたデータの信頼性の品質保証がなされる。統計解析を担当する統計家は計画段階から参画し、解析計画書を策定し解析実務を行い、膨大な量となる報告書作成に協力する。

新薬申請の場合は、このような臨床試験を通じて当該薬剤が安全性の点で許容可能であり、さらにもちろん患者の生存・QOLやこれらと直結していることが証明されている評価変数（エンドポイントと呼ばれる）の統計的処理を通じて有効であることの証明がなされて初めてデータが審査当局（日本の場合は厚生労働省）に提出される。そして、1年ほどの専門家による審査を経て、認可・発売となる。新薬1薬剤の開発費用はうなぎ上りであり、失敗例も含めれば1薬剤100-200億円以上、うち臨床試験の費用は数10億円以上となる。数千例を超えるような大規模試験の場合には数100億円の費用を要することもある。

2. わが国のevidence不足

さてEBMに戻ろう。わが国の臨床医学は、evidence作り、とくに臨床試験を通じたevidenceに実はほとんど寄与していない。その理由は、以下のような、臨床試験実施のためのインフラストラクチャを欠いていたことにある。

- (1) 医師研究者を支援するコーディネータ（Clinical Research Coordinator）
- (2) 研究計画書（プロトコル）を書くことのできる医師研究者とそのための教育
- (3) 生物統計家（試験統計家）あるいは生物統計学の教育
- (4) 効率的なデータマネジメント・システムと中立的なデータセンター
- (5) Peer reviewあるいは監査・査察によるデータの品質保証体制
- (6) 医師研究者主導型の臨床研究に対する法的規制
- (7) 結果を発信するプロフェッショナルであるMedical writerの人材と教育
- (8) 臨床試験の意義の理解と患者参加に対するインセンティブ

(9) 臨床試験に参加する医師研究者へのインセンティブ

(10) 研究資金とくに公的研究資金の供給と評価

「治験」とは、製薬会社がスポンサーとなり新薬申請あるいは適応拡大申請のために行われる臨床試験である。臨床試験には、製薬会社のスポンサーで市販後に行われる（通常は大規模な）市販後臨床試験や、公的研究費を用いて行われる医師研究者主導型の研究、更にはとくにこのような支援無しで行われる自主臨床試験が含まれる。

さて治験あるいは製薬会社スポンサーの市販後臨床試験に関しては、GCP (Good Clinical Practice) あるいはこれに準じた GPMPSP (Good Post Marketing Surveillance) としてその倫理性と科学性とを保証する仕組みができ上がっている。これは日・米・欧の三極の規制当局と製薬業界との国際合意 ICH-GCP (ICH は「国際ハーモナイゼーション」の略称) を受け、わが国では 1998 年から完全施行されている。GCP とはヒトを対象とする臨床試験の計画・実施・記録および報告に対する国際的な倫理的科学的品質基準であり、この遵守によって被験者の権利・安全性および福祉が世界医師会による「ヘルシンキ宣言」に基づく原則に沿った形で保護され、臨床試験データは信頼できるという公的な保証が与えられる。この GCP に応えるため、少なくとも製薬会社主導の試験に関しては、上記のインフラストラクチャのかなりの部分がわが国でもようやく整備された。CRC の主要施設への配置やこれを派遣する SMO (Site Management Organization) の設立、施設側の体制整備、モニタリングや監査による品質保証の定着、製薬会社の体制整備、そしてメディアを通じた患者リクルートや教育も開始された。生物統計学の認知もその現れであり、製薬会社や CRO には修士レベル以上の統計実務家が配置されるようになった。しかし、医師研究者主導の市販後の臨床試験に対するインフラストラクチャはまだ未整備である。標準治療に直結する evidence は実はこのような市販後の臨床試験によって生み出されることが多い。わが国で evidence の生れなかった事実、生物統計学を含むこのような試験実施のインフラストラクチャが未整備であった結果に他ならない。

なぜ統計学が不要であったかは、わが国医学界の臨床研究軽視と根は同一である。戦後渡米した日本の俊英英医学者は、言語・免許の壁から基礎医学研究の基盤

を日本に導入した。これは現在花を咲かせ、一方、臨床医学の研究基盤は形成されなかった。これまでの新薬開発・薬事制度がこれに拍車をかけた。類似薬を先例通りに開発すれば認可され、保険制度の中で利益を確保できたから。わが国の臨床試験の歴史は長い、残念ながら最近までの統計の応用はお作法であり、科学研究・技術評価の道具では必ずしもなかった。

3. 医薬品開発・標準治療確立のために行われる臨床試験

医薬品開発あるいは標準治療開発のために行われる臨床試験の流れを表 1 に示した。

健常人に対して行われ安全性を検討するのが第 I 相、患者に対して行われ、漸進的な有効性の検討がなされるのが前期第 II 相（海外では IIa と呼ばれる）である。この段階では、用量に関してランダム化したいくつかの群を設定する場合も、群別にあるいは個人内で用量を漸増するデザインの双方が存在する。試験薬の安全性が十分検討されていない段階であるので、後述する「盲検」（どの薬剤を用いているのか区別できなくする工夫）無しで試験が行われることもある。対象者数を増やし、有効性と安全性のバランスから臨床用量の選択を行うのが後期第 II 相 (IIb) である。通常は、有効成分を含まないプラセボを含んだ 3 から 5 群が設

表 1 臨床試験の進め方。

	通常の薬剤	抗がん剤
第 I 相	<ul style="list-style-type: none"> ● 健常人対象（専門施設） ● 単回、連投試験 ● 安全性検討、薬物動態 	<ul style="list-style-type: none"> ● 通常は患者対象 ● 複数スケジュール、増量 ● MTD・推奨用量決定
第 II 相	<ul style="list-style-type: none"> ● 患者対象、通常並行群 ● 有効性の検討、用法用量の決定 ● 安全性の検討 	<ul style="list-style-type: none"> ● 患者対象、通常単群 ● 腫瘍縮小による有効性の確認 ● 安全性（毒性）の検討
第 III 相	<ul style="list-style-type: none"> ● 患者対象、通常並行群 ● 標準治療との非劣性（優越性）か ● プラセボとの優越性の検証 	<ul style="list-style-type: none"> ● 日本では市販後 ● 標準治療との比較（通常組合せで）

MTD: Maximum tolerated dose

定され盲検下で試験が実施される。次の第 III 相では、他に有効な対照薬が存在しない場合にはプラセボ対照、広く使われている対照薬が存在する場合にはその薬剤を対照として、当該試験薬の有効性が検証される。副作用（国際標準では有害薬物反応）対策の点で患者のリスクが過大とならない限り、盲検下で試験が実施されるのが普通である。プラセボを対照とした場合には、試験薬の優越性を検証するように、すなわち群間差の検定が有意となる確率（検出力）を一定以上にしよう試験が設計される。通常、検出力は 80-90% に設定される。実薬が対照である場合には、この優越性か、試験薬がある程度以上劣っていないことを検証する「非劣性試験」が設計される。わが国の治験では、それぞれの相で対象となる患者数は、おおよそ、第 I 相で 20-30 人、前期第 II 相で 30-100 人、後期第 II 相で 100-400 人、第 III 相で 200-500 人程度である。

臨床試験において有効性を測定する指標・変数がエンドポイントである。例えば降圧剤開発のエンドポイントは血圧（の降下作用）、高脂血症なら総コレステロールか LDL コレステロール（の降下作用）がエンドポイントであり、これらを指標として新薬が開発されている。しかしこれらは、患者の立場に立った場合には、必ずしも生存や QOL に直結するとは限らない。降圧剤であれば、最終的な目標は血圧のコントロールを通じて脳卒中に代表される循環器系疾患を予防することである。抗高脂血症薬も主な目標は冠動脈疾患の予防である。これら最終的なエンドポイントを対象として、標準治療を確立するために行われる臨床試験は mortality・morbidity 試験と呼ばれるが、対象とした疾患の発症頻度が必ずしも高くないため、また多くのリスク因子がその発症を修飾し、後述の言葉で言えば誤差的バラツキが大きくなるため、数 1000 例ときには 10000 例を超える症例数を必要とする巨大研究となるのが普通である。当然、試験の計画から解析までには数年から 10 年を費す。有望な薬剤をできるだけ早く患者の手に届けるため、また化合物の特許を保護し製薬会社に新薬開発のインセンティブを与える観点から、このような mortality・morbidity 試験は市販後に、しばしば製薬会社とは独立に医師研究者主導型で行い、これに先立つ認可はこれらの最終的なエンドポイントと疫学的に高い相関を有するエンドポイント評価でなされることがある。この意味で、最終的なエンドポイントに替わるエンドポイントを代替エンドポイント (surrogate endpoint) と呼ぶ。抗癌剤開発に

表 2 最近 10 年の臨床試験の歴史。

1993.9	ソリブジン事件
1996.5	ICH-GCP (E6 ガイドライン) 国際合意
1996.6	薬事法改定
1997.3	答申 GCP, 省令 GCP 通知
1997.4	医薬品機構誕生
1997.7	医薬品・医療機器審査センター誕生
1998.4	GCP 完全実施
1998.8	ICH-E5 ガイドライン「海外臨床データ受け入れにおける人種要因差」通知
1998.11	ICH-E9 ガイドライン「臨床試験のための統計的原則」通知
2002.7	薬事法改定
2003.6	「医師主導の治験の実施の規準」通知 「臨床研究に関する倫理指針」通知

における腫瘍縮小も代替エンドポイントである（毒性が強く健常人への投与が行えない抗癌剤等では、臨床試験の組み立てがかなり異なっているがここでは省略しよう）。

さて、わが国の臨床試験を取り巻く環境と実施状況は、表 2 に示すように最近 10 年間で大変革を迎えた。その引き金は前述の国際ハーモナイゼーション (ICH) であり、1993 年 9 月から 11 月の間に 15 人が副作用で亡くなったソリブジン事件であった。ヘルペスに対して開発された治療薬ソリブジンが（頻用されていたが、当時承認されていたのはわが国のみの）5FU 系統の経口抗癌剤の代謝を阻害し、重篤な薬剤相互作用を引き起こしかねないことは基礎的検討から事前に認識されていた。しかも治験中には死亡例も発生していた。しかしながら、これらの事実は軽視され、認可とそれに引き続く活発な宣伝活動により一気にソリブジンが使われ犠牲者を頻出させた。治験の質と審査体制の問題が明らかとなった本事件が主なきっかけとなり審査体制が見直され、新設の医薬品審査センターに、生物統計家 2 人が初めての統計専門審査官として厚生労働省内部に迎えられた。

製薬会社主導であれ、医師研究者主導であれ、ICH-E9 ガイドライン「統計的原則」によって臨床試験に携わる試験統計家には「適切な資格と経験 (appropriately qualified and experienced)」が要求されるに至った。試験統計家の地位はわが国でも確立し、その要件と基本的な考え方も国際標準化されたといつてよい。民族差に関するガイドライン (E5 ガイドライン) により、海外データを新薬申請に公式に用いることも可能となった。現在の状況は、既存の海外データを利用して国内

の治験をいかにして「サボるか」から、国際的な同時承認を目指して臨床試験をいかに国際展開するか、日本人に対する標準的な治療をいかに確立するか、再生医療や遺伝子治療などの最先端医療の臨床試験をいかに迅速に行うか、DNP チップなど遺伝子情報も生かし個別化を目指した医療（テーラーメイド医療）の開発研究をいかに展開するかに移りつつある。臨床統計・生物統計の input はますます必要とされる状況が現出しつつある。

4. 臨床試験の目標と方法論の貢献

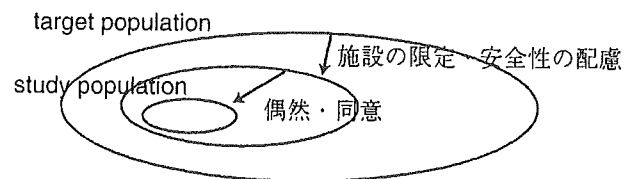
臨床試験計画時の目標は、統計的には次の3点にまとめられる。

1. **Clarity** : ランダム誤差を小さくし、研究の精度を向上させる
2. **Comparability** : 治療群間の比較可能性（研究の正確性）を向上させる
3. **Generalizability** : 結論の一般化可能性を向上させる

現状の我々の知識水準では発生理由を同定できず（あるいはその必要はなく）、ランダムとみなさざるをえない（あるいはランダムとみなして操作上は差し支えない）バラツキをランダム誤差と呼び、これを小さくすることが clarity の要求である。このためのもっとも強力な対策はサンプルサイズを増やすことであり、あるいは誤差の小さな反応変数をエンドポイントとすることである。予後に大きな影響を及ぼす予後因子を計画時から考慮し（ランダム化を層化して行う、あるいは後述の動的割付法の変数に用いる）、解析時に層別あるいは数学モデルにより調整することによっても研究の精度は向上する。測定やデータ管理における品質管理も重要な手段である。

Comparability の要求とは、系統誤差すなわちバイアスを減らすことに対応し、研究の正確性を向上させることと言い替えてもよい。もっとも強力な手段はランダム化であり、既知の因子については、層化要因あるいは動的割付法の変数として考慮する（当然解析時には、精度を上げるため調整の対象とすべきである）ことも重要な手段である。介入手段（通常は薬剤）の種類を医師・患者双方に対し識別不能とする2重盲検法は患者選択と管理・評価にわたって comparability を高く保つための手段である。なお、2重盲検試験で

ランダム化 randomization 内的妥当性の保証
無作為抽出 random sampling 外的妥当性の保証



無作為抽出：多くの統計理論において理論上重要な概念であるが、臨床試験においてはほとんど現実味をもたない。私見であるが、不用で誤解を招く概念。

Senn, S(1997), *Statistical Issues in Drug Development*, Wiley: glossary.

図2 ランダム化と無作為抽出。

は、データ管理や解析の担当者も盲検化されており、この意味で3重盲検と呼ぶほうがふさわしい。

リストの最後にある generalizability は、しばしば混同して用いられる無作為抽出 (random sampling) とランダム割付け (random allocation) を対比させることにより理解がしやすくなる。図2に示すように、臨床研究では、研究成果を適用する目標集団 (target population) と実際に研究が行われる対象集団 (study population) の間には数の上でも質の上でも大きな隔たりが生じる可能性が高い。まず、安全性の考慮から患者除外条件が設定され参加施設の選択がこれに続く。(例えば試験参加医師の外来担当日にたまたま来院したなど) ランダムに対象が選ばれる側面も多少はありうるが、研究に同意し参加する患者はランダム標本とは考えられない。すなわち、臨床試験研究においては(世論調査や社会調査などで統計手法の妥当性の根拠となる) 無作為標本やこれを保証する無作為抽出は存在しない。この、偏りを含み得る対象集団から妥当な結論を引き出すための方法がランダム化であり、この妥当性を internal validity と呼ぶ。比較試験結果の検定に際しては、無作為標本であることは必ずしも前提として必要ではなく、ランダム化によっても、検定の妥当性は保証される(なお、臨床試験においては、E9の翻訳において無作為化でなく「ランダム化」を標準用語として採択した。これは、後述するように単純な無作為化ではなく、多様な患者背景のバラツキを減ずるように、ランダムな要素を伴って割付けが行われる実態を強調するためである)。

したがって、厳密に言えば、得られた結論の適用範囲は当該の対象患者のみであり、この結論が目標集団まで外挿できるかどうか一般化可能性あるいは外的

妥当性 (external validity) の議論である。対象集団の背景分布が目標集団のそれと著しく異なる場合には一般化可能性は低いだろう。患者を性・年齢・疾患重症度などの背景要因によって層別し、これによって得られる部分集団 (層) を解析し治療効果の差が層毎に異なるなら、あるいは多施設共同試験では、施設内の治療効果差が施設毎に大きく異なるなら、一般化可能性は高いだろう。アメリカの薬品・食品の認可当局 FDA (Food Drug Administration) は、第 III 相試験を最低 2 つ行うことを原則として義務付けている。これも一般化可能性を高めるためであり、メタアナリシスも一般化可能性を検討する手段である。

5. バイアスの制御とランダム化

偏り (バイアス) とは何らかの原因により、結論が系統的に真実からずれる現象であり、臨床研究においては以下の 3 つが主な原因とされている。

- (1) 対象の選択
- (2) 試験の実施 (情報収集と評価)
- (3) 交絡 (Confounding)

交絡とは、患者反応に影響を及ぼす因子の分布が群間でアンバランスとなり、介入効果が正しく反応に反映しない現象である。重症の患者に薬剤が選択的に投与される結果、薬剤投与群が非投与群に劣る現象がその例である。臨床試験においては、上記 (2) に対しては盲検化によって、(1) と (3) とくに未知の因子による交絡に対してはランダム化による防止策が採用される。そして、臨床医学に対する統計学の最大の貢献はこのランダム化の導入であり、技術評価として 1930 年代に R.A. Fisher によって提唱された実験計画法とされている。新種や肥料を評価する農事試験を対象として Fisher が創案した実験計画法の基本的な考え方は、以下の概念にまとめられる。

1. バラツキの大きさを評価し、同時に推論の感度を向上させるための反復
2. 系統的なバラツキを小さくするための局所管理
3. 局所管理によっても除去しえない潜在的な、あるいは未知の要因によるバラツキを偶然誤差に転化し、確率論に基づく評価を可能とする、言い替えば平均的に均質で比較可能な群を作るためのランダム化

4. 多数の因子の影響を効率的に評価できるよう、それらを同時に変化させて割付けを行う多因子要因実験

Fisher の実験計画法は、1947 年に英国 Medical Research Council によって開始された結核患者に対するストレプトマイシンの評価に採用され、その成功を通じ、臨床試験においても方法論としての有効性が確立されたと考えられている。臨床試験において、1 はそのまま採用された。2 は農事試験においては圃場をブロックという区画に分割することによって達成されるが、臨床試験においては重要な予後因子や参加施設で患者を事前層別することによって行われる。同一の患者に異なる治療を行う (経口薬の場合であったら日を変えて、塗布薬であったら場所を変えて) クロスオーバー試験は、患者個人をブロックとした研究計画である。3 は、説明と理解の上での同意という農事試験には存在しない重大な倫理的問題を引き起こしたが、臨床試験にも採用され、4 は、癌や循環器の大規模な予防研究等において積極的に採用されるに至っている。

6. 臨床試験に対する生物統計家の貢献

臨床試験に携わる生物統計家の役割の第一は、これまでの知識・先行研究を受け、臨床家との討議を通じ、データ解析によって証明できる形に仮説を設定することにある。このためには何をもって試験治療を有効とするか、いわゆる研究のエンドポイントを設定することが前提となる。新しい分野であれば、このエンドポイントの信頼性・妥当性を事前に評価することも必要となる。ランダム化試験の場合に割付けの方法を検討すること、仮説を証明するために必要最小な症例数を設定すること、偏りのない、かつ感度の高い解析法を提案すること、より探索的な相においてはデータを適切に要約するモデルを検討することも統計家の役割であり、参加患者の利益を考慮し、不必要に試験を継続しないような中間解析の方法も事前に検討される。実際のデータ解析・報告書作成に協力することも、もちろん統計家の役割である。以下、統計家の課題をやや詳しく述べよう。

- (1) 証明すべき仮説は明確に、データ解析によって検証できる形に述べられているか。

検定によって新しい治療法が有意に優れることを示

す場合が多いだろうが、毒性が強い実施が困難な標準治療に対しては、試験治療の存在意義を非劣性（著しく劣っていない）の形で示せば十分な場合もある。この場合には閾値の設定と試験感度（assay sensitivity：本当に有効な治療が「有効」と結論されるかどうか）の確保が問題となる。後者は試験の品質管理に依存するところ大である。

現在、非劣性の検証は、臨床的に同等とみなせる差を試験治療に「上乘せ」して「明らかに劣る」という仮説を棄却し「同等以上」であることを示す検定に進じた方法と、試験治療と対照の効果の差（あるいは比）の信頼区間を計算し、これを上記の許容できる範囲と比較する方法とが用いられている。

(2) サンプルサイズの設定の根拠と方法は妥当か？

実は「必要サンプルサイズ」はきわめて曖昧で、おおよその目安を与えるにすぎない。結局は、当該分野である程度のコンセンサスを獲得して「臨床的に意味のある差」を設定し計算するのが実態である。十分なサンプルサイズが確保できないことが明らかならば、（必要なら海外との共同研究の形で）メタアナリシスを初めから考慮した計画とすることも考えられる。また、十分な事前情報が存在しない場合には、盲検化でこれまでのデータを中間解析し、必要サンプルサイズあるいは（長期追跡研究の場合であれば）追跡期間を延長することも行われる。最近では、独立データモニタリング委員会が開鍵した解析を行い、その結果に基づき症例数追加・追跡期間延長勧告を行う試験デザインの妥当性が議論されている。

(3) 割付け方法は妥当か？

盲検化が実施困難ながんの多施設共同研究においては、ファクシミリや電話さらには Web を用いた中央登録法はすでに常識化されている。盲検化が可能な（試験の）場合であっても、試験実施状況を迅速に把握し予後要因の偏りを防ぐ意味で、中央登録が行われる例が増えてきた。

割付けにおいて層別するとすればどのような要因を用いるか、それまでの患者の登録状況（背景要因の分布）に応じ偏りがより小さくなる方向に割付ける可能性を増やす「動的割付け法」を用いる場合には、どの因子を用いるか、施設をどう考慮するか、などが問題となる。層別要因はせいぜい 1-2、動的割付けの場合なら施設以外に 2-5 程度の因子を考慮することが可能

である。施設に関しては、予見性を少なくし、かつ治療群の偏りを防ぐために動的割付けの因子とするか、極端な偏りが生じないような調整法（偏りの大きさに応じて割付け率を変える「偏コイン法」など）が用いられている。

(4) 中止・脱落・治療不遵守（違反例）例の取扱いは妥当か？

研究の目的をよく理解し、また取扱いによって選択バイアスがどのように生じるかを予測すれば、症例の取扱いは多くの場合ほぼ自明である。例えば経口剤を用いた予防研究の場合であつたら、いわゆる intent-to-treat の方針に従って、有害薬物反応や患者の（途中での）投薬拒否例は分母から除くべきではないことは明らかである。有害薬物反応あるいは飲みにくさ自体が薬の負の効果であり、患者選択によって予後不良の可能性の高い患者が除かれるからである。一方、治療開始後の不適格判明については、その判定が割付けられた治療群や、治療結果に依存していないことが示されれば、解析から除くことも当然ありうる。現在では、有効性と安全性について解析対象集団を事前にプロトコルに定義するのが普通である。また、対象の取扱いによって結果がどのように変化するか「感度解析」を行うことを明記するプロトコルも増えている。

(5) 解析計画は妥当であるか、また解析結果の解釈は適切か？

検定のみならず、治療効果の推定を行うべきであることが近年強調されている。もちろん、その前提として、治療効果を単一のパラメータに縮約できるかどうかの検討が必要であり、推定結果の信頼性も信頼区間という形で示さねばならない。なお、予後に影響しない因子については、それが群間で偏りがあつたとしても、調整に用いる必要はない。逆に予後に強く影響する因子については偏りが存在しなくとも解析感度（精度）を上げる意味で調整に用いるべきであり、このことはプロトコルに事前に明記することが望まれる。調整の方法としては、層別と結果の併合（連続反応変数なら分散分析、0-1 あるいは順序反応変数あるいは Time-to-event なら Cochran-Mantel-Haenszel 流のアプローチ）あるいは回帰分析型の統計モデル（連続反応変数なら重回帰分析、0-1 反応変数ならロジスティック回帰分析、time-to-event であつたら Cox 回帰）を用いるという 2 つのアプローチが用いられる。前者はラン

ダム化に確率計算の基本を置く design-based な解析法であり、後者は説明変数と反応変数の間の関連、および反応変数のバラツキに確率モデルを想定することから、model-based な方法と呼ばれている。

繰り返し測定データに対する Generalized Estimating Equations, 混合効果モデル, 並べ替え検定など高度・コンピュータインテンシブな手法の利用も, 統計パッケージの普及もあってはもはや常識的である。

(6) 中間解析

予想以上の効果が早期から試験治療群に見られた場合,あるいは逆にこれ以上試験を継続しても試験治療の優越性(あるいは非劣性)を証明できないことが強く予想された場合に,試験を打ち切るための統計的な基準・方法が種々考案されている。代表的な方法が,事前に設定した回数の検定を多重性を考慮し行う群逐次検定(group sequential method)である。しかし,統計的基準以上に重要なことは,迅速に解析に必要なデータを収集・集計する統計センターを含むデータ管理体制であり,また,研究者とは独立な立場で評価を行うことのできる独立データモニタリング委員会(効果・安全性評価委員会)の存在である。データ固定の方法,試験途中での固定に適した調査票の設計も実務上は重要である。

臨床試験の統計的側面に関する参考文献

臨床試験の計画と統計解析に要求される水準は,近年きわめて高くなりつつあり,その状況を見るには最初の2つのガイドラインが適切であろう。1992年のわが国の統計解析ガイドラインは,2)の国際合意(ICH)ガイドラインを受け廃止された。臨床試験の統計的側面について全般については3,4)が,大規模試験の実施については5)が,がん臨床試験については7,8)が優れた教科書である。6)は入門として価値がある。最近の動向を知るには9)が参考となろう。サンプルサイズの設定については10)がまとまっている。11)から15)は最近の教科書で,11)はがんを中心とし倫理面の記述が多いことが特徴。13)は新薬開発を対象としているが,皮肉とウイットの効いた著述スタイルが面白い(日本語訳は困難)。12)はアメリカ最大のがん共同研究機構 SWOG の統計センター

のスタッフによるコンパクトな医師向けのテキスト。14)は臨床薬理全般の教科書であるが,医薬品開発と統計的側面にも詳しい。16)は専門家向けの包括的な参考書(辞典)である。17)は,わが国のがん共同研究機構である JCOG のプロトコルマニュアルであり,実務上きわめて詳細かつ有用である。

- 1) 薬審第 335 号 (1996 年 5 月 1 日):「治験の総括報告書の構成と内容に関するガイドライン (E3)」。
- 2) 薬審第 1047 号 (1998 年 11 月 30 日):「臨床試験のための統計的原則 (E9)」。
- 3) Pocock (著)・コントローラ委員会訳:『クリニカル・トライアル』,篠原出版,1989。
- 4) Friedman, LM., Furberg, C. and DeMets, D.: *Fundamentals of Clinical Trials*, Third ed. Springer, 1998。
- 5) Meinert, C.: *Clinical Trials*, Oxford Univ. Press, 1986。
- 6) 折笠秀樹:『臨床研究デザイン』,真興交易医書出版社,1994。
- 7) Leventhal, B. and Wittes, R. (著)・福島雅典・大橋靖雄(監訳):『がん臨床研究の方法』,メディカルブックサービス,1995。
- 8) Buyse, M. et al. ed.: *Cancer Clinical Trials*, Oxford Univ. Press., 1985。
- 9) Thall, P.: *Recent Advances in Clinical Trial: Design and Analysis*, Kluwer Academic Publications, 1995。
- 10) Machin, D. and Campbell, M.: *Statistical Tables for the Design of Clinical Trials*, Blackwell Scientific Publications, 1987。
- 11) Piantadosi, S.: *Clinical Trials, A Methodologic Perspectives*, Wiley, 1997。
- 12) Green, S., Benedetti, J. and Crowley, J.: *Clinical Trials in Oncology* second edition, Chapman and Hall, 2003。
- 13) Senn, S.: *Statistical Issues in Drug Development*, Wiley, 1997。
- 14) 日本臨床薬理学会編:『臨床薬理学 第二版』,医学書院,2003(とくに I 章 E. 医薬品の開発)。
- 15) 丹後俊郎:『無作為化比較試験』,朝倉書店,2003。
- 16) Redmond, C. and Colton, T. ed.: *Biostatistics in Clinical Trials*, Wiley, 2001. (Wiley Reference Series in Biostatistics)。
- 17) JCOG プロトコルマニュアル
<http://jcogweb.res.ncc.go.jp/>

(おおはし・やすお, 東京大学大学院医学系研究科,
NPO 法人・日本臨床研究支援ユニット理事長)