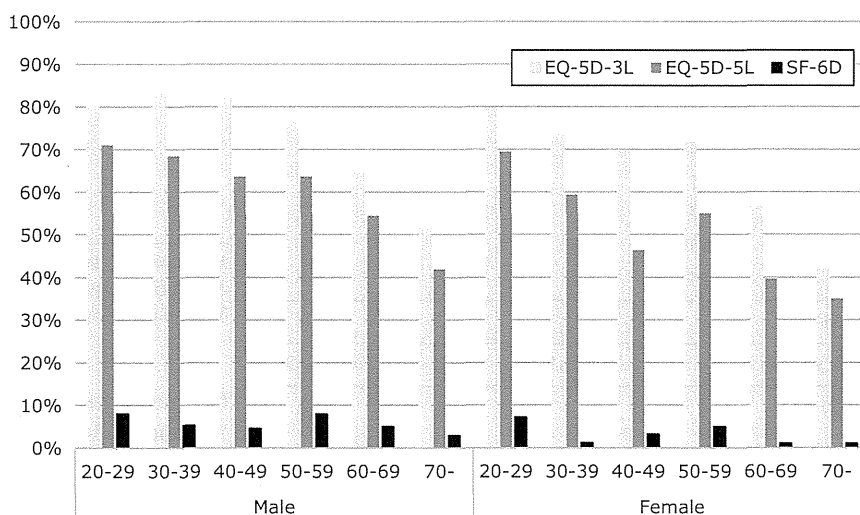


Fig. 2 Comparison of Japanese population norms with those of other countries. **a** EQ-5D-3L, **b** SF-6D

Fig. 3 Percentage of respondents reporting full health



The results are shown stratified according to sex and age category. The QOL scores were significantly reduced if the respondents were older than 60 years of age, female, had a lower income, or a shorter period of education. According to our results, a larger income was associated with a higher QOL score. The causal relation (whether poverty causes a poor health state or a poor health state is the cause of poverty) is unclear, but this finding may be useful for public health policies. This relation was observed in other countries. For example, in the USA [14], the QOL score as measured using the EQ-5D-3L was 0.81 for the poorest category (\leq USD 10,000), although it was 0.92 for the richest (\geq USD 75,000).

The percentage of reports of any health problem for the EQ-5D-5L is higher than that for EQ-5D-3L in almost all the sex and age categories. Some authors have pointed out that the EQ-5D-3L has a ceiling effect, which is defined as “the proportion of respondents scoring ‘no problems’ on any of the five dimensions” [45], because the instrument lacks enough sensitivity. A three-level questionnaire allows respondents with a slightly worsened health state to be reported as having a full health state. This is one example of how the ceiling effect problem has been improved by the revision of the EQ-5D-3L, resulting in the EQ-5D-5L. According to Table 2, the standard deviation of the QOL score measured using the EQ-5D-5L tended to be smaller than that measured using the EQ-5D-3L. This result may also arise from the increased number of levels, enabling respondents to choose intermediate levels.

Compared with the EQ-5D measures, the QOL score measured using the SF-6D was lower in the general population. A poor agreement between the EQ-5D and the SF-

Table 4 Percentage of respondents reporting any problem in each dimension

| (a) EQ-5D | | | | |
|-------------|--------|--------|--------|--------|
| Age (years) | Male | | Female | |
| | 3L (%) | 5L (%) | 3L (%) | 5L (%) |
| 20–29 | | | | |
| MO | 3 | 2 | 4 | 3 |
| SC | 1 | 2 | 2 | 1 |
| UA | 3 | 2 | 3 | 2 |
| PD | 10 | 20 | 13 | 22 |
| AD | 14 | 22 | 11 | 19 |
| 30–39 | | | | |
| MO | 1 | 3 | 2 | 3 |
| SC | 0 | 1 | 0 | 0 |
| UA | 1 | 1 | 3 | 3 |
| PD | 13 | 22 | 22 | 35 |
| AD | 9 | 24 | 10 | 17 |
| 40–49 | | | | |
| MO | 2 | 2 | 5 | 7 |
| SC | 1 | 1 | 1 | 1 |
| UA | 6 | 7 | 8 | 8 |
| PD | 13 | 28 | 23 | 43 |
| AD | 8 | 20 | 16 | 27 |
| 50–59 | | | | |
| MO | 6 | 6 | 6 | 8 |
| SC | 1 | 2 | 0 | 1 |
| UA | 8 | 9 | 5 | 6 |
| PD | 16 | 32 | 27 | 42 |
| AD | 13 | 18 | 11 | 16 |
| 60–69 | | | | |
| MO | 10 | 12 | 11 | 13 |
| SC | 4 | 4 | 4 | 6 |
| UA | 11 | 13 | 13 | 15 |
| PD | 33 | 43 | 37 | 56 |
| AD | 10 | 13 | 18 | 23 |
| ≥70 | | | | |
| MO | 24 | 34 | 32 | 37 |
| SC | 11 | 13 | 11 | 12 |
| UA | 27 | 25 | 26 | 27 |
| PD | 35 | 50 | 50 | 61 |
| AD | 15 | 19 | 12 | 20 |

(b) SF-6D

| Age (years) | Male (%) | Female (%) |
|-------------|----------|------------|
| 20–29 | | |
| PF | 26 | 32 |
| RL | 15 | 13 |
| SF | 28 | 22 |
| BP | 49 | 56 |
| MH | 74 | 76 |

Table 4 continued

| (b) SF-6D | | |
|-------------|----------|------------|
| Age (years) | Male (%) | Female (%) |
| VT | 85 | 84 |
| 30–39 | | |
| PF | 45 | 56 |
| RL | 11 | 19 |
| SF | 22 | 30 |
| BP | 55 | 64 |
| MH | 68 | 70 |
| VT | 88 | 94 |
| 40–49 | | |
| PF | 56 | 58 |
| RL | 17 | 28 |
| SF | 30 | 34 |
| BP | 58 | 69 |
| MH | 70 | 80 |
| VT | 91 | 90 |
| 50–59 | | |
| PF | 63 | 74 |
| RL | 23 | 20 |
| SF | 22 | 33 |
| BP | 55 | 67 |
| MH | 62 | 64 |
| VT | 85 | 88 |
| 60–69 | | |
| PF | 76 | 92 |
| RL | 32 | 43 |
| SF | 39 | 46 |
| BP | 73 | 84 |
| MH | 59 | 66 |
| VT | 86 | 87 |
| ≥70 | | |
| PF | 91 | 96 |
| RL | 39 | 48 |
| SF | 31 | 43 |
| BP | 71 | 81 |
| MH | 57 | 57 |
| VT | 85 | 85 |

MO mobility, SC self-care, UA usual activities, PD pain/discomfort, AD anxiety/depression, PF physical functioning, RL role limitation, SF social functioning, BP bodily pain, MH mental health, VT vitality

6D scores was observed, with a low ICC of 0.249 (EQ-5D-3L) and 0.234 (EQ-5D-5L). One cause seems to be clear, considering the percentages of respondents with full health as shown in Table 4. The percentage of people who chose no problem on the SF-6D was much lower than that for either EQ-5D measure. This result may be characteristic of the SF-6D and not only for the Japanese population. In

Table 5 Relation between QOL scores and diseases and symptoms

| | | EQ-5D-3L | EQ-5D-5L | SF-6D |
|---------------------------------|----------|----------|----------|----------|
| <i>(a) Diseases</i> | | | | |
| Intercept | | 0.9631 | 0.9618 | 0.7572 |
| Age (years) | | | | |
| 20–29 | 192 | – | – | – |
| 30–39 | 160 | 0.0134 | 0.0046 | –0.0077 |
| 40–49 | 189 | –0.0034 | –0.0079 | –0.0257 |
| 50–59 | 185 | 0.003 | –0.0016 | 0.0061 |
| 60–69 | 202 | –0.0232 | –0.0167 | –0.0296* |
| ≥70 | 214 | –0.0552* | –0.0384* | –0.0308* |
| Sex | | | | |
| Male | 556 | – | – | – |
| Female | 586 | –0.0185* | –0.0102 | –0.0229* |
| Diseases | | | | |
| No diseases | 608 | – | – | – |
| Diabetes | 42 | –0.0893* | –0.0898* | –0.0558* |
| Dyslipidemia | 24 | –0.0045 | –0.0192 | –0.0353 |
| Depression or mental diseases | 22 | –0.1934* | –0.1739* | –0.1845* |
| Eye diseases | 16 | –0.0677* | –0.0543* | –0.0420 |
| Ear disorder | 8 | –0.0390 | –0.0293 | –0.1088* |
| Hypertension | 94 | –0.0236 | –0.0181 | –0.0402* |
| Stroke | 8 | –0.2361* | –0.1675* | –0.1370* |
| Angina or myocardial infarction | 18 | –0.0707* | –0.0642* | –0.1054* |
| Cold | 7 | –0.0343 | –0.0356 | –0.1264* |
| Stomach and duodenal diseases | 9 | –0.0878 | –0.0823* | –0.0941* |
| Tooth disorder | 45 | –0.0207 | –0.0259 | –0.0494* |
| Rheumatoid arthritis | 9 | –0.1786* | –0.1835* | –0.0846* |
| Arthrosis | 15 | –0.1472* | –0.1173* | –0.0929* |
| Omalgia | 19 | –0.0260 | –0.0659* | –0.0856* |
| Lumbago | 37 | –0.1114* | –0.1026* | –0.0963* |
| | <i>n</i> | EQ-5D-3L | EQ-5D-5L | SF-6D |
| <i>(b) Symptoms</i> | | | | |
| Intercept | | 0.9912 | 0.9844 | 0.7893 |
| Age (years) | | | | |
| 20–29 | 192 | – | – | – |
| 30–39 | 160 | 0.0020 | –0.0013 | –0.0172 |
| 40–49 | 189 | –0.0104 | –0.0167 | –0.0326* |
| 50–59 | 185 | –0.0111 | –0.0157 | –0.0055 |
| 60–69 | 202 | –0.0409* | –0.0292* | –0.0378* |
| ≥70 | 214 | –0.0800* | –0.0664* | –0.0510* |
| Sex | | | | |
| Male | 556 | – | – | – |
| Female | 586 | –0.0173* | –0.0085 | –0.0216* |
| Symptoms | | | | |
| No symptoms | 588 | – | – | – |
| Having a fever | 6 | –0.1211* | –0.0890* | –0.0885 |
| Feeling sluggish | 8 | –0.1481* | –0.1382* | –0.2050* |
| Sleeplessness | 12 | –0.2012* | –0.1637* | –0.2078* |
| Feeling irritable | 14 | –0.0661 | –0.0915* | –0.1503* |

Table 5 continued

| | <i>n</i> | EQ-5D-3L | EQ-5D-5L | SF-6D |
|---------------------------------------|----------|----------|----------|----------|
| Headache | 25 | −0.1205* | −0.1102* | −0.1130* |
| Dizziness | 11 | −0.1152* | −0.1125* | −0.1569* |
| Blurred vision | 12 | 0.0220 | −0.0019 | −0.0738* |
| Having trouble seeing | 13 | −0.0778* | −0.0512 | −0.1457* |
| Buzzing | 9 | −0.0501 | −0.0719* | −0.0663 |
| Having trouble hearing | 9 | −0.1331* | −0.0641 | −0.1362* |
| Palpitation | 10 | −0.1300* | −0.1022* | −0.1430* |
| Cough and/or sputum | 24 | −0.0395 | −0.0264 | −0.0919* |
| Nasal congestion or mucus | 43 | −0.0509* | −0.0296 | −0.0638* |
| Feeling indigestion or heartburn | 10 | −0.0863* | −0.0828* | −0.0870* |
| Abdominal or stomach ache | 12 | −0.0846* | −0.0844* | −0.1176* |
| Dental pain | 10 | −0.0436 | −0.0113 | −0.0526 |
| Itching (such as eczema, tinea pedis) | 13 | −0.0543 | −0.0985* | −0.0534 |
| Stiff shoulders | 67 | −0.0482* | −0.0478* | −0.1007* |
| Back pain | 82 | −0.1142* | −0.1049* | −0.1303* |
| Arthritic pain | 40 | −0.1831* | −0.1247* | −0.1274* |
| Having trouble moving limbs | 10 | −0.2747* | −0.3618* | −0.2008* |
| Numbness of limbs | 15 | −0.1153* | −0.1092* | −0.1255* |
| Coldness of limbs | 11 | −0.1645* | −0.1566* | −0.1389* |
| Frequent urination | 10 | −0.0876* | −0.0846* | −0.1183* |

* *P* value < 0.05

Australia [33], the proportions of respondents in the 18- to 30-year age category who reported any problem in each dimension were as follows: 32 % for PF, 23 % for RL, 39 % for SF, 60 % for BP, 49 % for MH, and 94 % for VT. On the other hand, a Bland–Altman plot indicated that most outliers (an SF-6D score that was higher than the EQ-5D score) occurred at lower QOL scores. These tendencies were similar to those reported by Kontodimopoulos et al. [46] in Greece. Thus, the SF-6D may have a floor effect [47], i.e., the lowest QOL score of the SF-6D (0.292) is higher than that of the EQ-5D-5 (−0.025).

The Japanese population norms for the SF-6D seem to be lower than those for other countries, although that of EQ-5D-3L is similar to those of other countries (except Thailand). It is unclear whether this lower score is a result of the Japanese response pattern or a Japanese tariff for the SF-6D. According to these results, if the QOL score is used for economic evaluations, its interchangeability should be carefully considered [48–52], since the baseline scores of the general population differ between the Japanese EQ-5D and the SF-6D.

We analyzed the differences in the QOL scores between respondents with diseases/symptoms and those without diseases/symptoms by comparing the cross-sectional between-group MID of each measure. The anchor-based MID is more commonly measured longitudinally across

multiple time points, which is closer to the definition of MID. In the general population, repeated surveys are more difficult to perform than in clinical trials. Of note, our estimated score may not be the same as the intra-respondent MID. However, the between-group MID may be more useful when the results of between-group differences have been interpreted. Walters et al. [53] showed that the mean MID of the SF-6D was 0.041 and that of the EQ-5D-3L was 0.074 in a review of studies. In cancer patients, the MID of the EQ-5D was estimated to be 0.08 (UK score) and 0.06 (US score), and these values were anchored to the performance status and the FACT-G score [54]. According to a study examining post-traumatic stress disorder (PTSD), the MID was calculated as 0.05–0.08 (anchor-based method) and 0.04 to 0.10 (distribution method) [55]. Considering these scores, our MID is consistent with previous studies.

A limitation of this study was its relatively small sample size, compared with other studies to identify population norms. We think that the sample number was sufficient to estimate the population norms according to sex and age category, considering the interpretable and consistent results with previous studies in other countries. However, a larger number of subjects may enable a clearer relation between the QOL score and diseases/symptoms to be identified. Furthermore, analyses of the effects of diseases

with small prevalence could not be performed. Another limitation is the order in which the three instruments were presented to the respondents. As the order was fixed, and not randomized, the possible influence of the order on the results cannot be excluded based only on our data.

In conclusion, we demonstrated the following characteristics of three preference-based measures: (a) the Japanese population norms according to sex and age category, (b) the relation between QOL scores and socio-demographic factors, (c) the reliability of the three measures in the general Japanese population, (d) the percentage of reports of any problem, (e) the influence of diseases/symptoms on the QOL scores, and (f) the between-group MID. The respondents were randomly collected from all eight regions of Japan in a door-to-door survey, and the representativeness of the sample was considered to be good. The resulting information may be useful for calculating QALY in economic evaluations and research examining QOL score.

Acknowledgments We thank Makoto Kobayashi (CRECON MEDICAL ASSESSMENT INC.) for advice and support concerning this project. This work was supported by JSPS KAKENHI Grant No. 25871156.

Funding This study was funded by Grant-in-Aid for Scientific Research (KAKENHI) from Japan Society for the Promotion of Science (JSPS).

Compliance with ethical standards

Conflict of interest We have no conflict of interest to declare.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, 37(1), 53–72.
- Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care*, 35(11), 1095–1108.
- Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics*, 7(6), 490–502.
- Torrance, G. W., Furlong, W., Feeny, D., & Boyle, M. (1995). Multi-attribute preference functions. Health Utilities Index. *Pharmacoeconomics*, 7(6), 503–520.
- Torrance, G. W., Feeny, D. H., Furlong, W. J., Barr, R. D., Zhang, Y., & Wang, Q. (1996). Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Medical Care*, 34(7), 702–722.
- Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 health survey. *Journal of Clinical Epidemiology*, 51(11), 1115–1128.
- Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21(2), 271–292.
- Kharroubi, S. A., Brazier, J. E., Roberts, J., & O'Hagan, A. (2007). Modelling SF-6D health state preference data using a nonparametric Bayesian method. *Journal of Health Economics*, 26(3), 597–612. doi:10.1016/j.jhealeco.2006.09.002.
- Tsuchiya, A., Ikeda, S., Ikegami, N., Nishimura, S., Sakai, I., Fukuda, T., et al. (2002). Estimating an EQ-5D population value set: The case of Japan. *Health Economics*, 11(4), 341–353. doi:10.1002/heec.673.
- Ikeda, S., Shirowa, T., Igarashi, I., Noto, S., Fukuda, T., Saito, S., et al. (2015). Developing a Japanese version of the EQ-5D-5L value set. *Journal of the National Institute of Public Health*, 64(1), 47–55. doi:10.1007/s10198-013-0474-3. (in Japanese).
- Brazier, J. E., Fukuhara, S., Roberts, J., Kharroubi, S., Yamamoto, Y., Ikeda, S., et al. (2009). Estimating a preference-based index from the Japanese SF-36. *Journal of Clinical Epidemiology*, 62(12), 1323–1331. doi:10.1016/j.jclinepi.2009.01.022.
- Szende, A., Janssen, B., & Cabases, J. (Eds.). (2014). *Self-reported population health: An international perspective based on EQ-5D*. Netherlands: Dordrecht Springer.
- Kind, P., Dolan, P., Gudex, C., & Williams, A. (1998). Variations in population health status: Results from a United Kingdom national questionnaire survey. *BMJ*, 316(7133), 736–741.
- Luo, N., Johnson, J. A., Shaw, J. W., Feeny, D., & Coons, S. J. (2005). Self-reported health status of the general adult US population as assessed by the EQ-5D and Health Utilities Index. *Medical Care*, 43(11), 1078–1086.
- Fryback, D. G., Dunham, N. C., Palta, M., Hanmer, J., Buechner, J., Cherepanov, D., et al. (2007). US norms for six generic health-related quality-of-life indexes from the National Health Measurement Study. *Medical Care*, 45(12), 1162–1170. doi:10.1097/MLR.0b013e31814848f1.
- Konig, H. H., Bernert, S., Angermeyer, M. C., Matschinger, H., Martinez, M., Vilagut, G., et al. (2009). Comparison of population health status in six European countries: Results of a representative survey using the EQ-5D questionnaire. *Medical Care*, 47(2), 255–261. doi:10.1097/MLR.0b013e318184759e.
- Konig, H. H., Heider, D., Lehnert, T., Riedel-Heller, S. G., Angermeyer, M. C., Matschinger, H., et al. (2010). Health status of the advanced elderly in six European countries: Results from a representative survey using EQ-5D and SF-12. *Health and Quality of Life Outcomes*, 8, 143. doi:10.1186/1477-7525-8-143.
- Cunillera, O., Tresserras, R., Rajmil, L., Vilagut, G., Brugulat, P., Herdman, M., et al. (2010). Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey. *Quality of Life Research*, 19(6), 853–864. doi:10.1007/s11136-010-9639-z.
- Perneger, T. V., Combescure, C., & Courvoisier, D. S. (2010). General population reference values for the French version of the EuroQol EQ-5D health utility instrument. *Value Health*, 13(5), 631–635. doi:10.1111/j.1524-4733.2010.00727.x.
- Saarni, S. I., Harkanen, T., Sintonen, H., Suvisaari, J., Koskinen, S., Aromaa, A., et al. (2006). The impact of 29 chronic conditions on health-related quality of life: A general population survey in Finland using 15D and EQ-5D. *Quality of Life Research*, 15(8), 1403–1414. doi:10.1007/s11136-006-0020-1.
- Sorensen, J., Davidsen, M., Gudex, C., Pedersen, K. M., & Bronnum-Hansen, H. (2009). Danish EQ-5D population norms.

- Scandinavian Journal of Public Health*, 37(5), 467–474. doi:10.1177/1403494809105286.
22. Ferreira, L. N., Ferreira, P. L., Pereira, L. N., & Oppe, M. (2014). EQ-5D Portuguese population norms. *Quality of Life Research*, 23(2), 425–430. doi:10.1007/s11136-013-0488-4.
 23. Golicki, D., & Niewada, M. (2015). General population reference values for 3-level EQ-5D (EQ-5D-3L) questionnaire in Poland. *Polskie Archiwum Medycyny Wewnętrznej*, 125(1–2), 18–26.
 24. Johnson, J. A., & Pickard, A. S. (2000). Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. *Medical Care*, 38(1), 115–121.
 25. Clemens, S., Begum, N., Harper, C., Whitty, J. A., & Scuffham, P. A. (2014). A comparison of EQ-5D-3L population norms in Queensland, Australia, estimated using utility value sets from Australia, the UK and USA. *Quality of Life Research*, 23(8), 2375–2381. doi:10.1007/s11136-014-0676-x.
 26. Sun, S., Chen, J., Johannesson, M., Kind, P., Xu, L., Zhang, Y., et al. (2011). Population health status in China: EQ-5D results, by age, sex and socio-economic status, from the national health services survey 2008. *Quality of Life Research*, 20(3), 309–320. doi:10.1007/s11136-010-9762-x.
 27. Yu, S. T., Chang, H. Y., Yao, K. P., Lin, Y. H., & Hurng, B. S. (2015). Validity of EQ-5D in general population of Taiwan: Results of the 2009 national health interview and drug abuse survey of Taiwan. *Quality of Life Research*. doi:10.1007/s11136-015-0978-7.
 28. Abdin, E., Subramaniam, M., Vaingankar, J. A., Luo, N., & Chong, S. A. (2013). Measuring health-related quality of life among adults in Singapore: Population norms for the EQ-5D. *Quality of Life Research*, 22(10), 2983–2991. doi:10.1007/s11136-013-0405-x.
 29. Abdin, E., Subramaniam, M., Vaingankar, J. A., Luo, N., & Chong, S. A. (2014). Population norms for the EQ-5D index scores using Singapore preference weights. *Quality of Life Research*. doi:10.1007/s11136-014-0859-5.
 30. Kularatna, S., Whitty, J. A., Johnson, N. W., Jayasinghe, R., & Scuffham, P. A. (2014). EQ-5D-3L derived population norms for health related quality of life in Sri Lanka. *PLoS One*, 9(11), e108434. doi:10.1371/journal.pone.0108434.
 31. Menezes, R. M., Andrade, M. V., Noronha, K. V., & Kind, P. (2015). EQ-5D-3L as a health measure of Brazilian adult population. *Quality of Life Research*. doi:10.1007/s11136-015-0994-7.
 32. van den Berg, B. (2012). Sf-6d population norms. *Health Economics*, 21(12), 1508–1512. doi:10.1002/hec.1823.
 33. Norman, R., Church, J., van den Berg, B., & Goodall, S. (2013). Australian health-related quality of life population norms derived from the SF-6D. *Australian and New Zealand Journal of Public Health*, 37(1), 17–23. doi:10.1111/1753-6405.12005.
 34. Ferreira, P. L., Ferreira, L. N., & Pereira, L. N. (2015). SF-6D Portuguese population norms. *The European Journal of Health Economics*, 16(3), 235–241. doi:10.1007/s10198-014-0568-6.
 35. Ciconelli, R. M., Ferraz, M. B., Kowalski, S., Pinheiro, G. R. C., & Sato, E. I. (2015). Brazilian urban population norms derived from the health-related quality of life SF-6D. *Quality of Life Research*. doi:10.1007/s11136-015-0991-x.
 36. Fukuda, T., Shiroiwa, T., Ikeda, S., Igarashi, I., Akazawa, M., Ishida, H., et al. (2013). Guideline for economic evaluation of healthcare technologies in Japan. *Journal of the National Institute of Public Health*, 62(6), 625–640. doi:10.1007/s10198-013-0474-3. (in Japanese).
 37. Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, 20(10), 1727–1736. doi:10.1007/s11136-011-9903-x.
 38. Brazier, J. E., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care*, 42(9), 851–859.
 39. Ware, J. E, Jr, & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 30(6), 473–483.
 40. Fukuhara, S., Bito, S., Green, J., Hsiao, A., & Kurokawa, K. (1998). Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. *Journal of Clinical Epidemiology*, 51(11), 1037–1044.
 41. Fukuhara, S., Ware, J. E, Jr, Kosinski, M., Wada, S., & Gandek, B. (1998). Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. *Journal of Clinical Epidemiology*, 51(11), 1045–1053.
 42. Jenkinson, C., Stewart-Brown, S., Petersen, S., & Paice, C. (1999). Assessment of the SF-36 version 2 in the United Kingdom. *Journal of Epidemiology and Community Health*, 53(1), 46–50.
 43. Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 1(8476), 307–310.
 44. Eton, D. T., Cella, D., Yost, K. J., Yount, S. E., Peterman, A. H., Neuberg, D. S., et al. (2004). A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *Journal of Clinical Epidemiology*, 57(9), 898–910. doi:10.1016/j.jclinepi.2004.01.012.
 45. Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., et al. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: A multi-country study. *Quality of Life Research*, 22(7), 1717–1727. doi:10.1007/s11136-012-0322-4.
 46. Kontodimopoulos, N., Pappa, E., Papadopoulos, A. A., Tountas, Y., & Niakas, D. (2009). Comparing SF-6D and EQ-5D utilities across groups differing in health status. *Quality of Life Research*, 18(1), 87–97. doi:10.1007/s11136-008-9420-8.
 47. Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics*, 13(9), 873–884. doi:10.1002/hec.866.
 48. Ferreira, P. L., Ferreira, L. N., & Pereira, L. N. (2008). How consistent are health utility values? *Quality of Life Research*, 17(7), 1031–1042. doi:10.1007/s11136-008-9368-8.
 49. Whitehurst, D. G., & Bryan, S. (2011). Another study showing that two preference-based measures of health-related quality of life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? *Value Health*, 14(4), 531–538. doi:10.1016/j.jval.2010.09.002.
 50. De Smedt, D., Clays, E., Annemans, L., & De Bacquer, D. (2014). EQ-5D versus SF-12 in coronary patients: Are they interchangeable? *Value Health*, 17(1), 84–89. doi:10.1016/j.jval.2013.10.010.
 51. van Stel, H. F., & Buskens, E. (2006). Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease. *Health and Quality of Life Outcomes*, 4, 20. doi:10.1186/1477-7525-4-20.
 52. Yang, F., Lau, T., Lee, E., Vathsala, A., Chia, K. S., & Luo, N. (2014). Comparison of the preference-based EQ-5D-5L and SF-6D in patients with end-stage renal disease (ESRD). *The European Journal of Health Economics*. doi:10.1007/s10198-014-0664-7.
 53. Walters, S. J., & Brazier, J. E. (2005). Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research*, 14(6), 1523–1532.
 54. Pickard, A. S., Neary, M. P., & Cella, D. (2007). Estimation of minimally important differences in EQ-5D utility and VAS scores

- in cancer. *Health and Quality of Life Outcomes*, 5, 70. doi:10.1186/1477-7525-5-70.
55. Le, Q. A., Doctor, J. N., Zoellner, L. A., & Feeny, N. C. (2013). Minimal clinically important differences for the EQ-5D and QWB-SA in post-traumatic stress disorder (PTSD): Results from a doubly randomized preference trial (DRPT). *Health and Quality of Life Outcomes*, 11, 59. doi:10.1186/1477-7525-11-59.

A Review of Clinical Studies of Brand-name and Generic Drugs Used in Arrhythmia

Kaoru Ito¹⁾, Shunya Ikeda^{1) 2)}, Masaki Muto²⁾

Abstract

The objective of this study was to conduct a literature review comparing oral brand-name and generic drugs used for arrhythmia and to summarize the relevant clinical evidence.

Using the PubMed and Ichushi databases, we searched for articles comparing brand-name and generic drugs and categorized them according to the Vaughan Williams classification. For assessment, we divided the articles into affirmative and unfavorable groups according to the authors' positions concerning generic substitution. In addition, we evaluated the evidence levels of the articles.

Twenty articles were reviewed in this study, of which 14 were classified as affirmative and six as unfavorable. Of the affirmative articles, one was graded as evidence level I and six as evidence level II. Among the unfavorable articles, one was graded as evidence level II, four as evidence level V, and one as evidence level VI; no articles were graded as evidence level I. The affirmative articles included an evidence level I study report on drugs used for cardiovascular disease such as β -blockers; this evidence level demonstrates a significantly high level for articles pertaining to clinical efficacy and safety. Unfavorable articles tended to have lower evidence level than affirmative articles because many of these articles were case reports with a small number of subjects or descriptive studies, without details regarding the study methods and patients.

Keywords : Brand-name drugs, Generic drugs, Evidence level, Antiarrhythmic drugs

1. Introduction

In Japan, the national medical expenditure reached a record of 38.4 trillion yen in 2012 and is expected to greatly increase in the future (Ministry of Health, Labour and Welfare [MHLW], 2013). In October 2011, MHLW estimated the national health expenditure in 2025 to

be approximately 52 trillion yen, which is far beyond the 30-trillion-yen level in previous years (MHLW, 2011). In comparison with brand-name drugs, generic drugs are less expensive and equally safe. In this era of aging and increasing medical costs, the use of generic drugs could help control drug costs and support our national healthcare system. However, as of March 2013, the volume-based share of generic drugs remained 29.4% according to the "Recent trends of medical prescription fees" (MHLW, 2013).

According to the research on the validation of

¹⁾ Graduate School of Pharmacy, International University of Health and Welfare

²⁾ Graduate School of Health and Welfare Science, International University of Health and Welfare

medical service fee reforms published by the Central Social Insurance Medial Council, "drugs with strong pharmacological action and a narrow therapeutic index" was cited as a reason for medical practitioners avoiding the use of generic drugs for inpatients (Central Social Insurance Medial Council, 2013). Filing of generic drugs does not require clinical trials involving patients, and the equivalency of these drugs to their brand-name counterparts is only demonstrated using bioequivalent data obtained from healthy subjects. Therefore, some concerns remain such that generic drugs may cause unexpected blood level changes because of differences in the active ingredients, impurities, manufacturing methods, additives, or other factors compared to brand-name drugs.

The 8th Generic Medicines Quality Information Review noted that the publication of a study on amiodarone hydrochloride that questioned the efficacies of generic drugs had provoked domestic clinical sites to express their mistrust in generic substitutions (Generic Medicines Quality Information Review, 2012). Subsequently, the review approved the quality of an amiodarone hydrochloride tablet following a dissolution test (Generic Medicines Quality Information Review, 2012). However, in Japan, no evidence has demonstrated the efficacy and safety of generic antiarrhythmic drugs, which results in on-site concerns regarding the use of generic drugs to remain.

The objective of this study was to conduct a literature review comparing oral brand-name and generic drugs used for arrhythmia and to summarize the relevant clinical evidences.

2. Methods

1) Eligibility criteria

In this study, we covered a wide range of articles comprising not only specific study designs such as randomized controlled trials but also brief reports for comprehensive evaluation. The target drugs were oral formulations identical to antiarrhythmic drugs listed in the Vaughan Williams classification (VW classification), which is commonly used for the classification of antiarrhythmic drugs in Japan (2008 Joint Research Report, 2009). Parenteral products were excluded from this research. β -blockers approved for arrhythmia treatment were included.

2) Study selection

We conducted our search using the MEDLINE/PubMed and Igaku Chuo Zasshi/Ichushi-WEB databases (last research: January 2014). We used a combination of medical subject headings and keywords, and focused on the terms "generic drug" and "antiarrhythmic." The search formulas were as followed:

MEDLINE (databasc: PubMed)

- (1) "anti-arrhythmia agents" [MeSH Terms] OR (" anti-arrhythmia" [All Fields] AND "agents" [All Fields] OR "anti-arrhythmia agent" [All Fields] OR ("anti" [All Fields] AND "arrhythmic [All Fields]") OR "antiarrhythmic" [All Fields] OR "anti-arrhythmia agents" [Pharmacological Action])
- (2) "drug, generic" [MeSH Terms] OR ("drug [All Fields] AND "generic" [All Fields]) OR "generic drug" [All Fields] OR "generic" [All Fields]
- (3) (1) AND (2)

Igaku Chuo Zasshi (database: Ichushi-WEB)

- (1) 心臓血管作用剤/TH
- (2) 同種医薬品/TH OR 後発医薬品/AL
- (3) (1) AND (2)

We only included studies published in Japanese and English.

3) Data collection

Two authors (KI and SI) independently reviewed all titles and abstracts for eligibility using pre-defined criteria. All papers were considered as potentially eligible by one or two reviewers. Data were extracted into a pre-designed, structured Microsoft Excel[®] (Microsoft Corp., Redmond, WA, USA) form by one reviewer and appraised for completeness and accuracy by a second reviewer. Disagreements were resolved by discussion and, if necessary, the involvement of a third reviewer (MM).

4) Review procedure

(1) Article extraction

To extract clinical research articles about generic drugs from the retrieved articles, we excluded articles with the following characteristics: *in vitro* study, animal study, study of drugs other than oral administration products, comparison between brand-name drugs, no description of the generic drug, no description of clinical research, focus on cost-effectiveness and malpractice without a description of clinical research, and healthy subject study. When articles containing the same data were published in different journals by the same authors, the article that de-

scribed the study in greater detail was used. Case reports were not excluded.

(2) Grouping

Based on the contents of the articles extracted as described above, articles that met any of the following criteria listed below were classified into either affirmative article group or unfavorable article group.

- (1) A difference in pharmacokinetics was found between brand-name and generic drugs.
- (2) The brand-name drug was considered clinically more effective than the generic drug.
- (3) The generic drug caused adverse events more frequently than the brand-name drug.

(3) Data summary

We summarized the following data in tabular form: first author, year of publication, country where authors were located, study design, target formulation (brand-name and generic drug), target patients, observation period (or dose frequency), major evaluation items, and results. Afterwards, the evidence level of each article was determined based on the classification of evidence levels in the Handbook of Making Clinical Guidelines by Minds 2007 (Fukui *et al.*, 2007).

In addition, we investigated the distributions of affirmative and unfavorable articles and of evidence levels according to the authors' location. The Wilcoxon rank-sum test was conducted to statistically analyze the evidence level distributions in the affirmative and unfavorable article groups. A two-tailed *p*-value <0.05 was

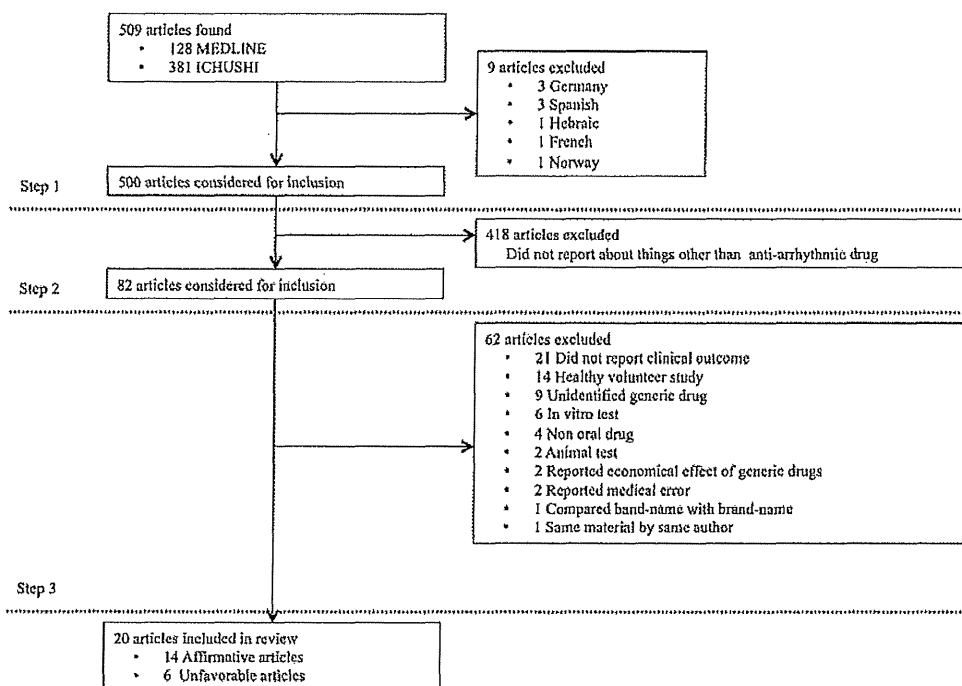


Figure 1 Study selection

considered significant. Data were processed using IBM®SPSS®Statistics software version 21 (SPSS, Inc., Chicago, IL, USA).

As this was a review, no particular consideration was given to ethical aspects.

3. Results

1) Target articles (Figure 1)

Our keyword search identified a total of 509 articles. After applying the preset exclusion criteria, 20 articles were extracted. Based on the contents, 14 articles were considered affirmative and six were considered unfavorable.

2) Summary of articles (Figure 2, Tables 1 and 2)

(1) First author

Through our MEDLINE search, we en-

sured that no articles by the same authors were included in either the affirmative or unfavorable group. However, an article by Kesselheim *et al.* (2008) cited five articles by authors that had articles in the affirmative group.

Through our ICHUSHI search, we observed that although all affirmative articles were written by different authors, two articles shared a coauthor. All unfavorable articles were reported by different authors.

(2) Year of publication

The oldest article was published in 1976. Among the included articles, none were published between 1976 and 1986. Two affirmative articles and two unfavorable articles were published between 1986 and 1989. Of the articles published in the

1990s, three were in the affirmative group and 1 in the unfavorable group. Five affirmative and three unfavorable articles were published in the 2000s, and three affirmative articles and no unfavorable articles were published in the 2010s.

(3) Countries where first authors were located

The majority of articles (n = eight; five affirmative and three unfavorable) were reported from the USA and four (three affirmative and one unfavorable) were reported from Japan. Four articles were reported from Europe, including from Israel, of which one unfavorable article was reported from the UK. One article was reported from Taiwan, the only Asian country other than Japan.

(4) Target drugs

Amiodarone hydrochloride (class III) was the most commonly evaluated drug with four reports. This was followed by atenolol (class II) with three reports. Thirteen of the 20 articles studied class II or III drugs according to the VW classification. No reports of class Ib drugs were identified.

(5) Study subjects

Seven of the 20 articles targeted patients with arrhythmia (five affirmative and two unfavorable). Six studies targeted patients with hypertension (five affirmative and one unfavorable). One article that included a questionnaire survey of arrhythmia specialists was included in the unfavorable group.

(6) Observation period

In the affirmative group, the longest observation period was nine years and two months in a prospective cohort study. Four studies did not specify an observation period. In the unfavorable group, five out of six articles did not specify an observation or treatment period.

(7) Major evaluation items

In the affirmative group, three articles evaluated clinical conditions, including subjective symptoms, electrocardiogram, heart rate, and blood pressure; three articles evaluated safety, including adverse events after switching to a generic drug from a brand-name drug; six articles evaluated pharmacokinetics, including the area under the blood concentration-time

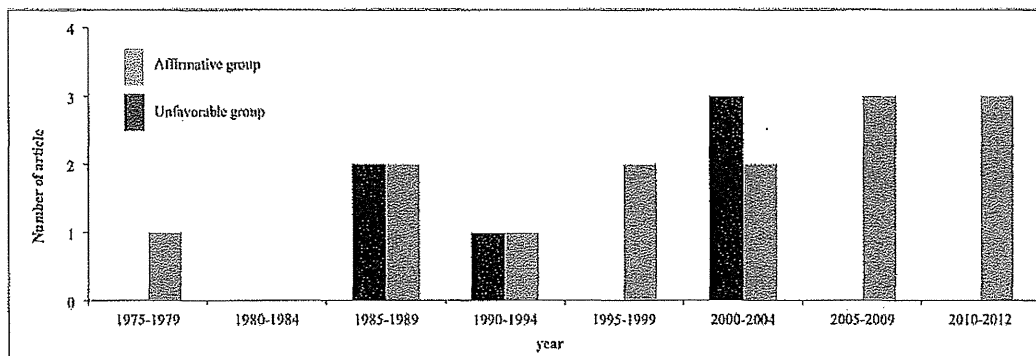


Figure 2 Year of publication

Table 1 Outline of the affirmative articles

| Ist Author | Published Year | Location | Study Design | Evidence Level | Target Formulation | Target Patient | Observation Period | Major Evaluation Item | Results |
|---------------|----------------|-------------|--------------|----------------|---|---|--------------------|--------------------------|--|
| Aoyanagi H | 2012 | Japan | CS | IVb | Amiodarone | 17 arrhythmia patients | Unclear | Clinical Symptoms/Safety | After switching to GE, new adverse events or noticeable changes in subjective symptoms, electrocardiogram findings, chest X-rays, and hepatic function were not observed. |
| Tsadok MA | 2011 | Canada | CS | IVa | Amiodarone | 60,220 arrhythmia patients with atrial fibrillation | 8 years | Safety | Cohort study using claims data. In 60,220 arrhythmia patients with atrial fibrillation, 2,905 patients took BD and 4,278 took GE. The incidence of thyroid gland deficiency was not significant. |
| Tanaka M | 2011 | Japan | CS | IVb | Amiodarone | 17 arrhythmia patients who had taken BD for >1 month and stability symptoms | Average 6.1 months | Clinical Symptoms/Safety | No endpoints were clinical differences or new adverse events. |
| Kesselheim AS | 2008 | US | SRMA | I | Metoprolol, Atenolol, Propranolol, Verapamil, Diltiazem, Procainamide | Unclear | Unclear | Clinical Symptoms | Identified 47 articles regarding 9 subclasses of cardiovascular medication, including β -blockers, to summarize clinical evidence comparing GE and BD. The aggregate effect size was -0.03 (95% confidence interval, -0.15 to 0.03) and did not indicate the superiority of BD over GE. |
| Tsuchishita Y | 2008 | Japan | Case Report | V | Disopyramide | 10 outpatients | Unclear | PK | In a TDM, no differences in the average blood concentrations were observed between BD and GE and no arrhythmic events were noted. |
| Ahrens W | 2007 | Germany | CS | IVb | Metoprolol | 49,673 patients with stable angina (health insurance companies in Germany) | 4 years | Safety | Claims database was used to compare hospitalization rates from stroke in BD and GE users via logistic regression analysis. The odds ratio of stroke (considering previous history of stroke and risk of thromboembolic disease) did not differ noticeably between the groups. (North Germany OR 1.04, Bremen: OR 1.06) |
| Amit G | 2004 | Israel | CS | IVb | Propafenone | 114 arrhythmia patients with atrial fibrillation | 18 months | Safety | Compared the incidence rates of AF, hospitalization, and emergency hospitalization in 114 arrhythmia patients with atrial fibrillation who used a BD for more than 18 months but switched to a GE for economic reasons. Although the incidence of emergency room visitation was higher among BD users than GE users, no differences in the incidence of electrical cardio version and/or pacemaker implantation were observed. |
| Sauro SC | 2002 | UK | CS | IVb | Amiodarone | 138 arrhythmia patients | 2 years | PK | TDM revealed no significant difference in the average blood concentrations of amiodarone and desethylamiodarone between BD and GE users. |
| Saseen JJ | 1997 | UK | RCT | II | Verapamil | 16 hypertension patients aged \geq 60 years | 7-15 days | PK | No significant differences in blood pressure, heart rate, ECG (P-R interval), or PK (AUC, Cl, Cmax, Tmax) between BD and GE. |
| Chiang HT | 1995 | Taiwan | RCT | II | Atenolol | 23 hypertension patients | 4 weeks | Clinical Symptoms | No noticeable differences in heart rate and blood pressure between BD and GE groups. |
| Rostock G | 1992 | Germany | RCT | II | Atenolol | 12 hypertension patients | Single dose | PK | AUC did not differ significantly between BD and GE. |
| Carter BL | 1989 | US | RCT | II | Propranolol | 15 hypertension patients | 12 weeks | Clinical Symptoms | No significant difference in blood pressure variation between BD and 2 types of GE. |
| Kasmer RJ | 1987 | US | RCT | II | Procainamide | 10 arrhythmia patients | Single dose | PK | No significant difference in blood procainamide concentration between BD and GE. |
| Saeterboek AM | 1976 | Netherlands | RCT | II | Quinidine | 24 MI patients | Unclear | PK | No significant difference in blood quinidine concentration between BD and GE. |

BD: brand-name drug, CS: comparative study, GE: generic drug, SRMA: systematic review and meta-analysis, TDM: therapeutic drug monitoring, AF: atrial fibrillation, RCT: randomized controlled trial study, AUC: area under the (blood concentration-time) curve, Cl: clearance, Cmax: maximum drug concentration, Tmax: maximum drug concentration, MI: myocardial infarction, OR: odds ratio

curve (AUC) and maximum blood concentration (Cmax) of the active ingredient and its biometabolites; and two articles evaluated both clinical conditions and

safety. In the unfavorable group, there were no articles evaluating clinical conditions. However, three articles addressed pharmacokinetics and safety.

Table 2 Outline of the unfavorable articles

| Ist Author | Publication Year | Location | Study Design | Evidence Level | Target Formulation | Target Patient | Observation Period | Major Evaluation Item | Results |
|--------------|------------------|----------|-----------------------|----------------|--------------------|--|--------------------|-----------------------|--|
| Yokoshima T | 2004 | Japan | Case Report | V | Atenolol | 1 hypertrophic cardiomyopathy patient | 3 months | Safety | Experienced drug-induced liver disease from the addition of GE after switching from BD to GE. |
| Pollak PT | 2001 | Canada | Case Report | V | Amiodarone | 4 arrhythmia patients | Unclear | PK | Increased blood amiodarone concentration after switching from BD to GE in 4 cases. |
| Reiffel JA | 2000 | US | Expert Opinion | VI | Amiodarone | 130 clinical arrhythmia specialists | Unclear | Safety | Conducted questionnaire survey to obtain more information about frequencies of adverse effects from GE. Thirty-three of 66 clinicians experienced some types of adverse effects with GE. |
| Carter BL | 1993 | US | RCT | II | Verapamil | 8 healthy volunteers and 8 aging patients with hypertension | Unclear | PK | Cmax and AUC levels were elevated in aging patient who took GE versus those using BD ($p < 0.02$). |
| Grubb BP | 1989 | US | Case Report | V | Procainamide | A 52 year-old patient with ventricular tachycardia (Caucasian man) | Unclear | PK | Reduced plasma procainamide and n-acetylprocainamide concentrations after switching from BD to GE. |
| Sanderson JH | 1986 | UK | Cross-sectional Study | IVb | Procainamide | 1768 hypertension patients (60-79 years) | Unclear | Safety | Report of a part of a RCT implemented by the author's group. The GE group had an elevated baseline relative to the BD group ($p < 0.0001$). |

BD: brand-name drug, GE: generic drug, RCT: randomized controlled trial study

(8) Results

Eleven of 14 affirmative articles reported that there were no statistically significant differences between brand-name and generic drugs. The other three articles concluded that there were no significant differences between the products based on their evaluation results, although no statistical analyses were performed. Two of the six unfavorable articles concluded that there were statistically significant differences between the brand-name and generic drugs.

3) Classification of the evidence levels in clinical studies (Table 3)

Twenty target articles were classified based on evidence levels from clinical studies. In the affirmative group, one article was classified as level I, and six articles were classified as level II. In the unfavorable group, no article was classified as level I, whereas one article was classified as level II, three were IVb, one was V, and

one was VI. As a result, the evidence levels were significantly higher in the affirmative group than in the unfavorable group ($p = 0.026$).

The affirmative article by Kesselheim *et al.* (2008) received the highest evidence level classification in this study. Although Kesselheim *et al.* (2008) was included in both an affirmative and unfavorable article, we considered it an affirmative article based on the final results. The authors reported the results of a meta-analysis of the clinical equivalency of collected articles in which clinical efficacies were compared between brand-name and generic drugs used to treat cardiovascular diseases. In their report, effect sizes were calculated using data from 47 collected articles. Although their analysis included β -blockers, which are used as antiarrhythmic agents, the aggregate effect size was -0.03 (95% confidence interval, -0.15 to 0.08) for eight drug classes, demonstrating clinical equivalency between brand-name and generic drugs.

Five of the seven articles classified as level II

Table 3 Evidence levels of clinical studies

| | I | II | III | IVa | IVb | V | VI | Total |
|-------------------|---|----|-----|-----|-----|---|----|-------|
| Affirmative group | 1 | 6 | 0 | 1 | 5 | 1 | 0 | 14 |
| Unfavorable group | 0 | 1 | 0 | 0 | 1 | 3 | 1 | 6 |
| Total | 1 | 7 | 0 | 1 | 6 | 4 | 1 | 20 |

($p = 0.026$ Wilcoxon rank-sum test)

included pharmacokinetics evaluations. Four articles from the USA were classified as level I or II, whereas all Japanese articles were classified as level V or VI.

4. Discussion

In this study, we conducted a systematic literature review to collect a broad range of articles describing clinical studies on generic antiarrhythmic drugs and subsequently evaluated the scientific basis of each article in terms of the evidence level. Among the collected articles, affirmative articles included studies that analyzed sufficiently large panels of patients, including a meta-analysis conducted using integrating data from multiple randomized controlled studies comparing brand-name and generic drugs for cardiovascular diseases as well as case-control studies using receipt databases. Case reports on only a few patients and descriptive studies comprised majority of the identified unfavorable articles; these reports generally lacked detailed descriptions of study methods and patient information. A comparison of the evidence levels of the two article groups indicated a lower level for the unfavorable article group than for the affirmative article group.

Based on their contents, the 20 analyzed articles were divided according to clinical symptoms, pharmacokinetics, and safety. As a result, nine and eight articles were categorized as pharmacokinetics (of which two also evaluated

clinical symptoms in parallel) and safety, respectively. However, only five articles were categorized as clinical symptoms (of which two also evaluated pharmacokinetics in parallel). One likely reason for this imbalance is that a large proportion of the collected articles concerned amiodarone. Antiarrhythmic drugs have narrow blood concentration ranges and often require the use of an optimal blood level with large modifications to the standard administration method depending on individual patient conditions. Particularly for amiodarone, adverse and occasionally fatal reactions occur frequently during long-term oral treatment. Therefore, generic forms of amiodarone, which are manufactured differently, might have drawn attention. In addition, the comparison of a smaller number of articles categorized as clinical symptoms versus those categorized as pharmacokinetics or safety suggests that clinicians are most concerned about the "safety" of generic drugs.

Of the articles investigated in this study, the majority was published between 2000 and 2004, another finding that was attributed to amiodarone. In the USA, the first generic version of amiodarone was released in 1998 (FDA orange book). Two years thereafter, Reiffel and Kowey (2000) reported the results of a questionnaire survey of 130 arrhythmia specialists in the USA, which included questions regarding adverse events with amiodarone. This paper did not contain detailed information about the questionnaire

survey method or patient information pertinent to the adverse events. During the following year, Pollak (2001) reported that the blood levels of amiodarone and des-amiodarone (active metabolite of amiodarone) clearly varied among four patients who switched from a brand-name to a generic form of amiodarone. However, the article also failed to describe certain details, such as the study method and patient background, and did not give useful information about the relationship between generic drugs and patient background factors.

The articles on safety included affirmative articles that reported comparative studies with brand-name drugs and unfavorable articles comprising case reports. Tsadok *et al.* (2011) and Ahrens *et al.* (2007) comparatively assessed the incidence of adverse effects using claims databases from their respective countries. These observation studies might have been conducted in response to specific concerns of medical staff on generic drugs.

In the evidence level evaluation, the affirmative articles were found to have a significantly higher evidence level compared to the unfavorable articles ($p=0.026$). Kesselheim *et al.* (2008), whose article achieved the highest evidence level among the articles evaluated in this study, conducted a meta-analysis on the clinical effects of brand-name and generic drugs used to treat cardiovascular diseases (e.g., β -blockers and calcium blockers) that had been reported in articles published between 1984 and 2008, and demonstrated the clinical equivalence of brand-name and generic drugs. On the other hand, three out of six unfavorable articles were case reports or expert opinions with evidence levels only as high as V. The cases found in the case

reports included those in which the disease itself was uncommon, those where the disease was not uncommon but featured a rare symptom or course, cases in which special, uncommonly performed therapies were effective, and reports of adverse events. Meanwhile, case reports may involve information bias (i.e., the reported treatment might work only for a specific type of patient) and "chance" factors (i.e., the impact of an accidental treatment or event). Therefore, it is difficult to judge whether the therapeutic effects of the reported treatment are superior compared to existing treatments based on case reports.

Of the 20 articles analyzed in the present study, four were case reports. These included three unfavorable reports of adverse effects due to additives not found in brand-name drugs (Yokoshima *et al.*, 2004) and changes in the blood drug concentration that were not observed with brand-name drugs (Reiffel and Kowey, 2000; Grubb, 1989). These case reports were likely published with the intent to report events that have never been encountered with brand-name drugs. Regardless of the evidence level, such safety-related information is beneficial when, for example, a medical professional selects and uses a particular product from among many generic drugs or dispenses a generic drug as an alternative to the brand-name drug.

However, the magnitude and causality of risk cannot be evaluated from the results of a single study. Fact-based and evidence-based information should be collected and examined to remove concerns and suspicions that medical professionals and patients have and to allow informed use of generic drugs. To achieve this,

surveys and studies that use claims database information (e.g., disease name, diagnosis period, sex, age, and prescription) to evaluate the efficacy and safety of generic drugs will be very beneficial. In Canada, Tsadok *et al.* (2011) conducted a cohort study of claims data that contained information from 60,220 patients and reported no significant difference between brand-name drugs and generic drugs in terms of the incidence of thyroid dysfunction due to amiodarone (hazard ratio, 0.97; 95% confidence interval, 0.87-1.08). In Germany, Ahrens *et al.* (2007) assessed the rate of hospitalization due to cerebral infarction among patients using brand-name and generic drugs. In addition, they conducted analyses that considered the history of cerebral embolism and risk of embolism onset using claims databases from three German medical insurance companies and subsequently reported a lack of differences between brand-name drugs and generic drugs. Similar evaluations of generic drugs using receipt databases are being conducted in overseas countries and these epidemiological studies might also be necessary in Japan.

Study limitations

In this study, the articles were categorized based on content as affirmative or unfavorable according to our independently developed criteria and were evaluated in terms of the evidence level according to the study design. The classification method used for the affirmative and unfavorable articles in this study was not generalized and/or unambiguous. In addition, we did not distinguish articles by indication because only active ingredients belonging to the VW classification were of interest in this study. For example, articles on studies of hypertensive pa-

tients were included in studies of antiarrhythmic drugs such as atenolol and the evaluation was therefore not limited to antiarrhythmic drugs.

It should be noted that generic drug promotion policies were not considered in the present study. We omitted this consideration out of an expectation that these policies were unlikely to have any appreciable effect on clinical studies of generic drugs because systems and regulations, based on the Drug Price Competition and Patent Term Restoration Act (informally known as the Hatch-Waxman Act) (USA) and the Action Program for the Promotion of Safe Use of Generic Drugs (Japan), were mainly for the early introduction of generic drugs or environmental improvements for quality, supply systems, and information provision systems.

In general, an "unfavorable result" is less likely to be published compared to a "positive result," which is often referred to as "publication bias" (Easterbrook *et al.*, 1991). In the case of generic drugs, "unfavorable results" or reports of "no significant difference" from brand-name drugs appear more likely to be published compared to a "non-significant difference." In addition, this study included a case report describing a rare condition never encountered with brand-name drugs. Therefore, it is possible that a simple comparison of the evidence levels of all study designs leads to an evaluation that is unfavorable for brand-name drugs.

In the future to minimize the effect of bias, analyses should include, for example, the scoring of a target number of patients and the presence or absence of randomization in observation studies. It should also include observational studies comparing the efficacy and the safety of generic

drugs with the original medical supplies from a national claims database.

Despite the above-described study limitations, this study is of academic significance because no previous studies have comprehensively collected clinical studies comparing brand-name and generic oral antiarrhythmic drugs available in Japan and overseas and evaluated these studies based on the evidence level.

5. Conclusions

Through a systematic review, we comprehensively collected articles reporting clinical studies of generic antiarrhythmic drugs and evaluated the quality of the medical findings described therein according to the evidence level. Our results indicated that the unfavorable articles tended to have lower evidence levels compared with affirmative articles.

Acknowledgments

We are deeply grateful to Associate Professor Ataru Igarashi, Department of Drug Policy & Management, Graduate School of Pharmaceutical Sciences, The University of Tokyo, whose enormous support and insightful comments were invaluable for preparation of this article.

A part of this research was presented at the 5th Congress of Japan Society of Generic Medicine, (June 19, 2011).

Kaoru Ito is a graduate student of the Graduate School of pharmacy, International University of Health and Welfare and an employee of the medical affairs division, Janssen Pharmaceutical K.K. This study was carried out at previous employment (Sandoz K.K.). The content of this article reflects the personal view

of the authors only.

Reference

- Ahrens W, Hagemeyer C, Mühlbauer B *et al.* (2007) "Hospitalization Rates of Generic Metoprolol Compared with the Original Beta-blocker in an Epidemiological Database Study," *Pharmacoepidemiology and Drug Safety*. 16 (12) : 1298-1307.
- Amit G, Rosen A, Wagshal AB *et al.* (2004) "Efficacy of Substituting Innovator Propafenone for Its Generic Formulation Inpatients with Atrial Fibrillation," *The American Journal of cardiology*. 93 (12) : 1558-1560.
- Aoyanagi H, Okishige K, Miwa N *et al.* (2012) "The Evaluation of Efficacy and Safety after Switched from Brand Name Drug to Generic Drug of Amiodarone Hydrochloride Tablet," *Therapeutic Research*. 33 (1) : 103-108, in Japanese.
- Carter BL, Gersema LM, Williams GO *et al.* (1989) "Once-daily Propranolol for Hypertension : A Comparison of Regular-release, Long-acting, and generic formulations," *Pharmacotherapy*. 9 (1) : 17-22.
- Carter BL, Noyes MA and Demmler RW (1993) "Differences in Serum Concentrations of and Responses to Generic Verapamil in the Elderly," *Pharmacotherapy*. 13 (4) : 359-368.
- Central Social Insurance Medical Council (2013) "Summarize the Results of Investigation Usage Survey of Generic Drug after Revised the Medical and Technical Service Payment," in Japanese
(<http://www.mhlw.go.jp/file/06-Seisakujouhou-10800000-Iseikyoku/0000058316.pdf>) Accessed 2013/11/11.
- Chiang HT, Hou ZY, Lee DK *et al.* (1995) "A Comparison of Antihypertensive Effects between Two Formulations of Atenolol: Tenolol and Tenormin," *Chinese Medical Journal (Taipei)*. 55 (5) : 366-370.
- Easterbrook PJ, Berlin JA, Gopalan R *et al.* (1991) "Publication Bias in Clinical Research," *Lancet*. 337 (8746) : 867-872.
- Fukui T, Yoshida M and Yamaguchi N (2007) "Handbook of Making Clinical Guideline by Minds 2007," Igakushoin Co., Ltd., in Japanese.
- Generic Medicines Quality Information Review (2012) "The Summary of Meeting Minutes of the 8th Generic Medicines Quality Information Review meeting in February 22, 2012," in Japanese.
(http://www.info.pmda.go.jp/generic/generic_20120222.html) Accessed 2013/4/26.
- Generic Medicines Quality Information Review (2012) "The Summary of Meeting Minutes of the 9th Generic Medicines Quality Information Review meet-

- ing in September 26, 2012," in Japanese.
 (<http://www.info.pmda.go.jp/generic/generic_20120926.html> Accessed 2013/4/26.
- Grubb BP (1989) "Recurrence of Ventricular Tachycardia after Conversion from Proprietary to Generic Procainamide," *The American Journal of Cardiology*. 63 (20) : 1532-1533.
- Heart Electric Society Arts and Sciences Advisory Committee in Japan (2007) "Practices Series of Cardiovascular Pharmacological Treatment: How to Use the Amiodarone for Arrhythmia Treatment Revised Edition One," Life Medicom Co., Ltd.
- JCS Joint Working Groups for Guidelines for Diagnosis and Treatment of Cardiovascular Diseases (2008) "Guidelines for Drug Treatment of Arrhythmia (2009 revised edition)"
- Kasmer RJ, Nara AR, Green JA *et al.* (1987) "Comparable Steady-state Bioavailability between Two Preparations of Conventional-release Procainamide Hydrochloride," *Drug Intelligence and Clinical Pharmacy*. 21 (2) : 183-186.
- Kesselheim AS, Misono AS, Lee JL *et al.* (2008) "Clinical Equivalence of Generic and Brand-name Drugs Used in Cardiovascular Disease: A Systematic Review and Meta-analysis," *Journal of the American Medical Association*. 300 (21) : 2514-2526.
- Ministry of Health, Labour and Welfare (2011) "The Point of the Future Prospects the Medical Cost and Financial Effect Test Calculation in October 25, 2010"
 (<<http://www.mhlw.go.jp/stf/shingi/2r9852000000uhlp-att/2r9852000000uhq0.pdf>> Accessed 2014/08/16.
- Ministry of Health Labour and Welfare (2013) "Recent Trend of Medical Cost (approximate medical cost) March 2013," in Japanese.
 (<http://www.mhlw.go.jp/topics/medias/month/13/dl/gaisan_iryuu_data_03.pdf> Accessed 2013/11/11.
- Ministry of Health Labour and Welfare (2013) "Recent Trend of Medical Prescription Fees (for Electronic Processing) March 2013," in Japanese.
 (<<http://www.mhlw.go.jp/topics/medias/c-med/2013/03/>> Accessed 2013/11/11.
- Pollak PT (2001) "Altered Metabolite Concentrations with Amiodarone Generic Substitution Cannot be Observed without Monitoring," *Canadian Journal of Cardiology*. 17 (11) : 1159-1163.
- Reiffel JA and Kowey PR (2000) "Generic Antiarrhythmics are not Therapeutically Equivalent for the Treatment of Tachyarrhythmias," *The American Journal of Cardiology*. 85 (9) : 1151-1153.
- Rostock G, Vogel I, Stulich M and Günzel R (1992) "Bioequivalence of a new Atenolol Formulation," *International Journal of Clinical Pharmacology Therapy Toxicology*. 30 (11) : 482-483.
- Sanderson JH and Lewis JA (1986) "Differences in Side-effect Incidence in Patients on Proprietary and Generic Propranolol," *Lancet*. 1 (8487) : 967-968.
- Saseen JJ, Porter JA, Barnette DJ *et al.* (1997) "Postabsorption Concentration Peaks with Brand-name and Generic Verapamil: A Double-blind, Crossover Study in Elderly Hypertensive Patients," *The Journal of Clinical Pharmacology*. 37 (6) : 526-534.
- Sauro SC, DeCarolis DD, Pierpont GL *et al.* (2002) "Comparison of Plasma Concentrations for Two Amiodarone Products," *The Annals of Pharmacotherapy*. 36 (11) : 1682-1685.
- Soeterboek AM, Van Thiel M and Eng C (1976) "Serum Quinidine levels after Chronic Administration of Four Different Quinidine Formulations," *Journal of International Medical Research*. 4 (6) : 393-401.
- Tanaka M, Miyake S and Kurata A (2011) "The Evaluation of Efficacy and Safety after Switched from Brand Name Drug to Generic Drug of Amiodarone Hydrochloride Tablet," *Medical Consultation and New Remedies*. 48 (9) : 978-983, in Japanese.
- Tsadok MA, Jackevicius CA, Rahme E *et al.* (2011) "Amiodarone-induced Thyroid Dysfunction: Brand-name Versus Generic Formulations," *Medical Association Journal*. 183 (12) : E817-E823.
- Tsuchishita Y, Nunokawa S, Ito R *et al.* (2008) "Clinical Evaluations of Generic Drug: Comparison of Blood Levels of Disopyramide," *Japanese Journal of Generic Medicines*. 2 (Suppl. 6) 2, in Japanese.
- Yokoshima T, Tokunaga Y, Sato M *et al.* (2004) "A Case Report of Drug Hepatopathy was Caused by An Additive in Generic Drug ALCENOL," *Journal of Iwate Prefectural Hospital Association*. 43 (2) : 167-171, in Japanese.

(Received October 29, 2014, Accepted November 11, 2015)

Correspondence: Kaoru Ito
 kaoruitol@gmail.com

経口抗不整脈薬の先発医薬品と ジェネリック医薬品の 臨床比較研究の文献レビュー

伊藤かおる¹⁾, 池田 俊也¹⁾²⁾, 武藤 正樹²⁾

本研究では経口抗不整脈薬の先発医薬品とジェネリック医薬品を比較した臨床試験を網羅的に把握し、その内容のレビューと研究デザインをもとにエビデンスレベルについて評価した。文献は、Pubmedと医学中央雑誌を検索し、Vaughan Williamsの抗不整脈薬の分類表に表記されている薬剤を対象とした。また、ジェネリック医薬品に対する著者の記載内容から肯定的文献と否定的文献に分けて評価を行った。さらに収集した文献のエビデンスレベルを評価した。結果、20文献が今回の調査対象となった。内訳は肯定的文献が14文献、否定的文献が6文献だった。肯定的文献にはβブロッカーを含む循環器領域の治療薬を対象にしたエビデンスレベルIに評価される研究があるなど、臨床効果や安全性を評価した文献のエビデンスレベルが有意に高いことが明らかになった。否定的文献は、数症例を対象にした症例報告や記述研究による報告が多く、研究方法や患者の詳細な情報について記述がないものもあったことから、ジェネリック医薬品に対して否定的な文献の方が肯定的な文献よりもエビデンスレベルが低いと判断された。

キーワード 先発医薬品, ジェネリック医薬品, 臨床的同等性, エビデンスレベル, 抗不整脈薬

¹⁾ 国際医療福祉大学大学院薬学研究科

²⁾ 国際医療福祉大学大学院医療福祉学研究科