

## GER1000 QUANTITATIVE REASONING

### TUTORIAL 3

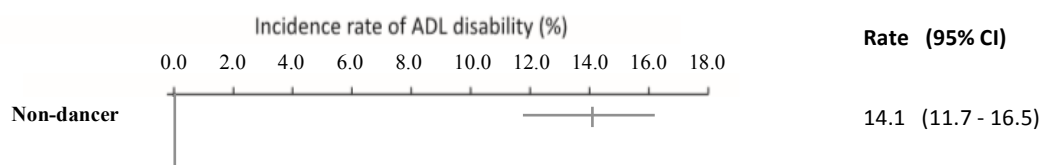
*Please work on the problems before coming to class. In class, you will engage in group work.*

#### Question 1

In 2018, the Scandinavian Journal of Medicine & Science in Sports published a study on whether dancing could reduce the risk of activities of daily living (ADL) disability for older women (70-84 years old). The article states “The target area of this study was the Itabashi ward, located in the northwest area of 23 special wards of Tokyo.”

The study used a probability sample from the Itabashi resident registry. For simplicity, we assume that a simple random sample was used. 16,883 older women from Itabashi were sampled and invited to enroll in the study. 15,880 women did not respond or refused to participate. Through interviews, 107 respondents were categorized as dancers, and the remaining 896 were non-dancers. ADL disability was measured through physical assessment. For parts (a)-(c), please ignore the non-response issue, i.e., treat the sample size as 896.

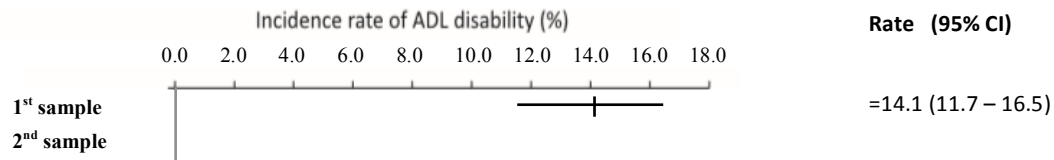
(a) The figure below shows a 95% confidence interval constructed from data on the non-dancers.



A friend looks at this figure and says: “There is a 95% chance that the rate of ADL disability among the Itabashi older women who are non-dancers is between 11.7% and 16.5%.” Is there anything wrong with the conclusion?

(b) After getting your advice, your friend says “I am 95% confident that the rate of ADL disability is between 11.7% and 16.5% among all the Itabashi older women who are non-dancers.” Interpret this statement using repeated sampling.

(c) Assume a second simple random sample is taken which is 9 times larger, i.e., its size is  $896 \times 9 = 8,064$ . It turns out the rate of ADL disability in the second sample is also 14.1%. Sketch the CI based on the second sample in the space below. You need to use the following fact: if a sample size increases by a factor of  $n$  (multiplied by  $n$ ), then the width of the confidence interval decreases by roughly a factor of the square root of  $n$  (divided by  $\sqrt{n}$ ). Based on the CIs, which of the two samples gives stronger support to the belief that the population rate of ADL disability is not 12.0%?



(d) Now we take the non-response into account. Is the interpretation of the confidence interval in (b) valid?

(e) The study reported: “Dancing was significantly and independently associated with a lower incidence of ADL disability”. More specifically, dancing seems to cut the risk of ADL disability by about 4 times. What subpopulation of the target population can the finding be generalised to?