Works Citing Bendel and Hua on Natural Fecundability
A Literature Review on the Origin of a Falsified Chart Used in High School Education in Japan

TANAKA Sigeto

Abstract

This paper reports the results of a literature review on “An Estimate of the Natural Fecundability Ratio Curve” by Bendel and Hua (1978, Social Biology 25). The estimation of this work was the origin of a falsified chart on women’s age-fertility profile that was featured in a high school health education material published in 2015 by the government of Japan. The author searched citation databases and collected 23 works citing the study. A review of the 23 works showed that biases and unreliability of the Bendel-Hua estimation had been pointed repeatedly. The results imply that the chart would be inappropriate for educational use, even if it were not falsified. Both the Japanese government and academics are responsible for the inappropriate chart being used without a comprehensive literature review to insure the reliability of scientific knowledge.

1. Background

In August 2015, the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) distributed a side-reader on the subject of health education in high schools, titled “健康な生活を送るために” [To Lead a Healthy Life], to all high schools in Japan (MEXT 2015a). On page 40 of the side-reader is a chart (Figure 1 (a)) that shows the relationship between a woman’s age and her likelihood of pregnancy, with a curve manipulated such that it peaks at the age of 22 years and then drops rapidly (Fassbender 2016).i

This chart was not newly created for this high school side-reader. The influential obstetrician and gynecologist Yoshimura Yasunori (吉村泰典) — a former chairperson of the executive board of the Japan Society of Obstetrics and Gynecology (2007–2011) and the Japan Society for Reproductive Medicine (2010–2014), who currently serves as a special advisor to the Cabinet (2013 to present) — had used the chart on his website, ii as lecture material, iii and in an online movie iv for public information prepared by the government (Tanaka 2016a; NDL 2016).

The findings presented in the chart are rooted in Bendel and Hua’s (1978) estimation of fecundability (the monthly probability of conception for a viable woman) based on the age-specific marital fertility rates (ASMFR) of American Hutterites (a natural fertility population with a high level of fertility) in the 1950–1960s (from Sheps 1965: Table 2), measured alongside data regarding the
length of time between marriage and first conception from a survey conducted in Taiwan (Jain 1969).

Figure 2 (a) shows the ASMFR profile with a three-year moving average, calculated from the Hutterite women’s data (Sheps 1965: Table 2) classified by age at the time of marriage, which was the original data used for Bendel and Hua’s estimation. ASMFR exhibit a common pattern, where rates are at a high of at least 50% during the “newlywed” period and decrease as the time from the marriage increases. The data indicates that, as long as they were in the early stages of their marital life, Hutterite women in their 30s exhibited the same levels of fertility as those in their early 20s.

Bendel and Hua designed their estimation of fecundability for women aged 25 and over using data from Sheps (1965: Table 2), after they removed data representing those women who married in their late 20s or after (Figure 2 (b)). That is, their estimation was based only on the two dotted lines in Figure 2 (a). Some parameters for Bendel and Hua’s estimation were extrapolated from other studies, such as the probability of miscarriage and stillbirth, duration of the unsusceptible period after childbirth, and sterility rate. Meanwhile, fecundability for women aged 16–24 was directly calculated from data on the duration of time between marriage and first childbirth, taken from the Taiwanese survey (Jain 1969). The resulting age-fecundability curve was a product of the combination of these two series of estimates (Figure 2 (c)). As a result of the above-mentioned removal of data representing those women who married in their late 20s or after, the right half of the age-fecundability curve exhibited a sharp decline as age advanced, a phenomenon that Bendel and Hua (1978: 217) described as “approximately linear.”

The curve was manipulated in a subsequent study by Wood (1989: Fig. 2.7), so that it hit its peak at the age of 22 (Figure 2 (d)). O’Connor et al. (1998: Fig. 3) then inaccurately reproduced this curve (Figure 2 (e)). Yoshimura (2013) drew the curve so that it hit its peak at the age of 22 and then declined linearly (Figure 2 (f)). Yoshimura then used this chart as evidence for public information and political activities, as we have mentioned above. A modified version of Yoshimura’s chart (Figure 1 (a)) was eventually featured in high schools around the country.

Within two days of MEXT publishing the side-reader, suspicion arose on the Internet that the chart was based on false information (Takahashi 2015). MEXT soon acknowledged that the chart contained errors. On September 2, 2015, MEXT issued a separate set of errata (MEXT 2015b), in which the curve of the chart was corrected to the one traced from O’Connor et al. (1998: Fig. 3). Afterwards, MEXT revised the online PDF version of its data on women’s fertility (MEXT 2015c). Figure 1 (b) is a copy of the chart from the revised PDF file. The Japan Society for Reproductive Medicine (JSRM) published a statement by the chairperson of the executive board (Irahara 2015) to justify the MEXT’s decision.

However, this revised chart (Figure 1 (b)) still contains two errors: (1) as we have seen, the original estimation by Bendel and Hua contains a number of methodological defects, and (2) the curve has been modified twice without proper scientific grounds after Bendel and Hua’s original publication,
and the current curve is not faithful to their original estimates (Figure 2 (c)). This paper addresses the first issue raised here—have problems with Bendel and Hua’s original estimation been recognized in the field of fecundability research?

2. Aim and method

In the following sections, we will examine works citing Bendel and Hua to explore what problems have been detected and how they have been evaluated in the field of fecundability research.

I searched databases to make a comprehensive list of works that cite Bendel and Hua. The search was conducted on February 3, 2016. Web of Sciences yielded 13 hits, Scopus yielded 7 hits, and Google Scholar yielded 17 hits, although one of these (Martin and Wu 1998) did not include Bendel and Hua in its reference list. As the results from these online database searches overlapped, the total number of retrieved works was 19. I located four additional works (Brewis 1992; Golden and Millman 1993; Weinstein et al. 1993; Wood 1994) through other means, so that I ended with a final total of 23 works to examine (see Appendix).

Note that the dissertation by Bendel (1978) contains similar findings to Bendel and Hua (1978). I was unable to locate any works citing Bendel’s dissertation.

3. Results

3.1. Estimates for Hutterite women aged 25 and over

A review of literature reveals two critical shortcomings in Bendel and Hua’s estimation of fecundability for Hutterite women aged 25 and over: (1) confusion between the effects of aging and the duration of marital life (James 1979; Wood 1989; Wood 1994), and (2) underestimation of the probability of sterility (James 1979).

Referring to Sheps (1965), James (1979: 333) points out that the Hutterite fertility data used by Bendel and Hua may have overestimated the effects of aging. A couple’s fertility may diminish as the time after marriage passes due to decreasing coital frequency, yet Bendel and Hua failed to distinguish between this effect and the effect of aging per se. James argues that only a small portion of the decline in fertility for women under 40 could be a result of aging itself.

Wood (1989: 77; 1994: 296, 318) advocates a similar point. He argues that the effect of the duration of marital life mediates the decrease of coital frequency. However, Wood (1989, 1994) does not mention Sheps in his argument, and it remains unclear whether he was aware of the defects in Bendel and Hua’s handling of data.

James (1979: 331) also discusses problems with the validity of an extrapolated parameter for
sterility rate (the parameter $s$ in the statistical model Bendel and Hua employed). Based on certain facts about Hutterite fertility, James argues that $s$ would be higher than the level that Bendel and Hua postulated. If so, Bendel and Hua’s figures may underestimate fecundability for those in their 40s.

My research found no additional criticisms of the other parameters Bendel and Hua extrapolated, such as the probability of miscarriage and the duration of the unsusceptible period after childbirth.

3.2. Estimates for Taiwanese women aged 16–24

Regarding the estimation of fecundability of Taiwanese women aged 16–24, Ellison (1994: 258) poses a question about the assumption that increasing fertility in the teenage females is a function of female reproductive physiology. Pennington and Harpending (1988: 304) argue that observed fertility might increase in the late teens due to the fact that teenage women are becoming newly fecund in that period. Bendel and Hua may have underestimated the degree to which women’s fecundability would be higher in their mid-teens because they used the fertility data averaged for both fecund and infecund women. These challenges are related to theoretical interpretations of the estimation, rather than to its methodological issues.

3.3. Other problems

Brewis (1992: 57) notes another theoretical problem: the notion of fecundability relies on an operational definition to denote the probability of conception within a month, so it cannot approximate fecundity, which denotes the ability to conceive in medical or biological term. This comment may be related to Ellison’s question mentioned above (Ellison 1994: 258).

Salo (1979) criticizes Bendel and Hua’s approach in gathering parameters from different populations. Salo points out that it would be difficult to guarantee the quality and comparability of all data from a heterogeneous sample of measurement results, and that “A model is only as good as its poorest input” (Salo 1979: 292).

A number of works quote Bendel and Hua’s study as a benchmark by which they evaluate the results of their own research. These works typically focus on establishing a rough correspondence between the shapes of their curves—often described as an inverse (or inverted) U-shape (Wood et al. 1994: 421; Strassmann 1997: 125; Strassmann and Warner 1998: 175). They do not rely on an exact comparison of fecundability at each age, nor do they highlight differences between specific details of their curves, with the exception of Wood and Weinstein (1988: 102), who report “our curve is somewhat flatter between ages 20 and 30.” Even the peak age might sometimes be inconsistent, as several authors wrote that fecundability hit its peak in the late 20s (Wood et al. 1994: 421; Dunson and Zhou 2000: 1057) or 30s (Williams 2003: 8) while citing Bendel and Hua.
4. Discussion

I found no work explicitly criticizing Bendel and Hua’s practice of removing data representing those women who married in their late 20s or after. However, some of them warn about confusion between the effects of aging and the duration of marital life, which may be a consequence of the data removal. In the year following the publication of Bendel and Hua’s research, James (1979) pointed out that their estimated fecundability would likely overemphasize the influence of age on declining fecundability. More recently, Wood (1989, 1994) addressed the same problem. In addition, there was another problem also raised by James about the underestimation of the probability of sterility.

As there has been no counter argument to the concerns raised by James and Wood, it seems reasonable to conclude that their criticism of bias in Bendel and Hua’s estimates has been accepted. Here it is worth noting that Bendel and Hua’s results are typically discussed in a favorable light with regard to overall shape of the curve. To be sure, the inverse U-shaped curve drawn by Bendel and Hua’s estimates is also an accepted theory of age-fecundability profiles. However, this applies only to a rough sketch of that curve, and the precise figure of the age-fecundability profile remains unknown. There is no consensus about the precise time when fecundability peaks in a woman’s life, nor about how rapidly it declines after that peak.

This study’s literature survey found no work containing a full explanation of the data and the methods used to replicate Bendel and Hua’s results. Without directly referring to Bendel and Hua or to Bendel’s dissertation, it would be impossible to know the exact process used to determine their estimation. Because I exclusively collected works from citation databases, you can expect that I succeeded in examining all works discussing methodological questions on Bendel and Hua, to the extent that relevant works were included in the databases.

This study reveals a consensus that Bendel and Hua’s estimates are biased and unreliable. In spite of the fact that this view is accepted by fecundability researchers, MEXT and JSRM have insisted to use the chart in school education, as we have seen. This practice is unscientific and exemplifies a misguided attitude toward education. MEXT and JSRM clearly recommended and approved the chart without taking the necessary steps to conduct a comprehensive literature review. This indicates that neither the government nor the designated academic association is adequately qualified to insure reliable communication of scientific knowledge.

Appendix

Extracts from the collected works (in order of publication date):

- The marks [S] [G] and [W] indicate the database in which the work was found (S: Scopus; G:
James (1979) [G, W] poses questions about the probability of sterility extrapolated into BH’s estimation: “Bearing in mind the very large number of children borne to Hutterite women, one might wonder whether these women reach menopause unusually early” (p. 331). It also argues that “the Hutterite data presented by Sheps (1965) ...... shows that Hutterite cumulative duration-specific birth rates over the first ten years of marriage are almost identical for women married at ages 25–29, and at ages 20–24, and at ages less than 20” (p. 333).

Salo (1979) [G] criticizes the method BH employed: “An available solution to the problem is to complete the input data required with estimates which have been obtained from prior studies on the populations of interest or ‘indeed from the data which do not even pertain exactly to this population’...... And there seems to be an expanding tendency to accept it among the workers in this area (c.f., Bendel and Chang-i Hua 1978). Too often, however, the biometric input achieved in this way consists of a heterogeneous sample of measurement results from computations with no evident historical, cultural or geographical comparability” (p. 292).


Sarma (1985) [G] briefly introduces BH’s data as an earlier study (p. 1).

Pennington and Harpending (1988) [W] cites BH in its literature review section about the pattern of fertility: “Many models show fecundity to be low at menarche, peak sharply during the mid-twenties, and then gradually decline to zero prior to menopause (......Bendel and Hua, 1978)” (p. 304). “Although the contribution of fecundity to fertility patterns is difficult to assess without methods for reliably assessing when a female enters the susceptible period, (Bendel and Hua, 1978), once a female has demonstrated her fecundity by giving birth to her first offspring, fecundity is probably at its maximum and only appears to increase subsequently due to averaging over a cohort with women who have not yet reached reproductive maturity” (p. 304).

Wood and Weinstein (1988) [S, W] cites BH in its literature review section: “Not surprisingly, then, there is empirical evidence that both total and effective fecundability vary systematically with age, rising from low levels after menarche to a broad peak during the mid-reproductive years and then falling again in the years preceding menopause."16 It appears to be no firm consensus about the relative contributions of the various factors determining fecundability to these age patterns."17 It also compares results of calculation with BH: “Even more gratifying, the detailed shape of the curve is quite similar to that constructed by Bendel and Hua from a variety of sources, save that our curve is somewhat flatter between ages 20 and 30” (p. 102).

Wood (1989) [G, W] adopts BH’s Fig. 1 with some manipulation (p. 77). It also argues the BH’s results include the effect of duration of marital life, as well as the age of the wife and the husband (p. 77).

Weinstein et al. (1990) [S, W] cites BH in its literature review: “It is now well established that effective fecundability varies in a systematic fashion with the age of the female partner, rising rapidly to a peak during her early 20s and then declining slowly to zero at about the time of menopause." However, there is a remarkable lack of agreement about the cause of these changes, in particular whether they are attributable primarily to changes in coital frequency, or to changes in the female’s physiological ability to begin and maintain a pregnancy." (pp. 447–448; Reference numbers 3 and 4 include BH).

Kinoshita (1990) [G, W] introduces BH as a study emphasizing the importance of behavioral determinants of fertility, especially coital frequency (pp. 330–331).
Brewis (1992) quotes a clause from BH: “A woman’s fecundability declines immediately after the first half of her twenties and ... that the decline is approximately linear all the way to the zero level near her menopause (Bendel and Hua 1978:217)” (p. 57). It also makes the criticism that “The assumption that fecundability approximates fecundity is not a realistic one, but it has been made” (p. 57).


Weinstein et al. (1993) cites BH in a paragraph introducing the history of fecundability research: “It is now well established that apparent fecundability varies in a systematic fashion with the age of the female partner, rising rapidly to a peak during the early 20s and then declining slowly to zero at about the time of menopause (Bendel and Hua, 1978). However, there is a remarkable lack of agreement about the cause of these changes, in particular whether they are attributable primarily to changes in coital frequency or to changes in the female’s physiological ability to begin and maintain a pregnancy (cf. Bendel and Hua, 1978......)” (p. 209–210).

Wood (1994) adopts BH’s Fig. 1 with some manipulation (p. 322). It also argues that the BH’s results include the effect of the duration of marital life, as well as the age of the wife and the husband (p. 296).

Wood et al. (1994) [S, G, W] compares the results of calculation with BH: “In agreement with earlier studies (Bendel and Hua 1978......), mean effective fecundability often follows an inverted U-shaped curve, rising to a peak in the late twenties and declining steadily thereafter” (p. 421).

Ellison (1994) [G, W] introduces BH as an earlier study: “As noted previously, little attention has been given to the separate question of increasing marital natural fertility in the first decade after menarche, although this is usually also assumed to be a function of female reproductive physiology (15, 65, 99, 154)” (p. 258. The reference number 15 denotes BH).


Strassmann (1997) [G, W] refers BH as one of the benchmarks used to evaluate data reported in the paper: “This result is consistent with reports that fecundability has an inverse U-shaped relationship with age (Bendel and Hua 1978......)” (p. 125).

Strassmann and Warner (1998) [S, G, W] refers to BH as one of the benchmarks used to evaluate the results of a statistical analysis: “Consistent with previous studies (Bendel and Hua, 1978......), fecundability had an inverse U-shaped relationship with age” (p. 175).

Smits et al. (1998) [S, G, W] cites BH to mention the low fecundability of women in their teens (p. 3522).

Dunson and Zhou (2000) [S, G, W] refers to BH as an information source used to determine the selection of variables in constructing a statistical model: “Mean effective fecundability based on live births has been observed ...... to rise to a peak in the late 20s and then decline steadily (Bendel and Hua 1978......)” (p. 1057).

Williams (2003) [G] cites BH as an earlier study: “Early studies on age-specific fertility suggested that rates increase with age and attain peak values in the 30s (......Bendel and Hua, 1978)” (p. 8).

Holman et al. (2006) [G] introduces BH as an earlier study about decomposing the factors involved in the process by which fecundability changes with age (p. 187).

Snopkowski and Kaplan (2014) [S, G, W] introduces BH as one of the sources from which their data were extrapolated (p. 327).
References


[MEXT 2015b] 文部科学省, 2015, 「高校 1年生用啓発教材『健康な生活を送るために』（平成27年度版）の訂正」 (2015年9月2日公表, 2015年11月30日取得,


【Sarma 1985】Sarma, R. S. S., 1985, “Parity Specific Birth Probabilities by Duration of Marriage” (CDC Working Paper 11), Cairo: Cairo Demographic Centre.


Acknowledgment

I would like to thank Editage (www.editage.jp) for English language editing.

Notes

1 The chart also exhibits problems of false labeling, concealing of data source, and inappropriate explanation. This paper, however, concentrates on the problem of unscientific data manipulation.

ii The earliest version of the chart was in an online article (Yoshimura 2013) dated June 25, 2013, published on the website of the Yoshimura Bioethic Institute, represented by Yoshimura himself. In the article, Yoshimura quoted the chart as demonstrating that women’s biological capacity to conceive declines in their 20s. He states that if we put the capacity to conceive as 1.0 at age 22, it will be lower than 0.6 at age 30, and will be about 0.3 at age 40. Similar charts appear on the same website, on two articles dated August 11 and November 15, 2014.


iv Japan’s Ministry of Health, Labour and Welfare (MHLW 2014) made a 12-minute movie to explain medical facts about pregnancy and infertility and the policies of the government related to those matters. It was made available to the public on the website YouTube, but has been unavailable since May 28, 2016. Yoshimura appeared on the movie to explain medical matters as an expert. The chart in question was presented in the latter half of the movie, as evidence indicating the suitable period for a woman to get pregnant. Yoshimura described the chart as showing that a woman’s capacity to conceive declines with age because of the reduction in the number of egg cells, and deterioration in their quality (Tanaka 2016b).

v Figure 2 (a) shows three-year moving averages, starting from the second year of each column in Table 2 of Sheps (1965: 68).

vi Note that there is discrepancy in the horizontal axis scale: compare Figure 1 (b) and Figure 2 (e).

vii The phrases “apparent fecundability” and “effective fecundability” seem to be occasionally conflated in the extracted texts. However, they are not underlined because it is unclear whether they were mistakenly used interchangeably, or whether they expressed different meanings. See Wood (1994: 280).
Figure 1. Charts on the high-school side-reader

(a) The first version: August 21, 2015 (MEXT 2015a: 40) (b) The revised version: September 30, 2015 (MEXT 2015c: 40)

Figure 2. The process of the transformation of the data