Hepatitis B and the Case of the Missing Women

Citation

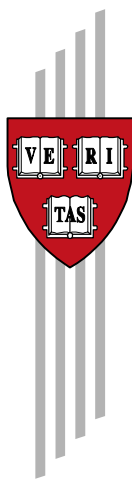## Published Version

Permanent link

Terms of Use

Share Your Story

Hepatitis B and the
Case of the Missing Women

Emily Oster

CID Graduate Student and Postdoctoral Fellow
Working Paper No. 7, March 2005

Working Papers

Center for International Development
at Harvard University

Hepatitis B and the Case of the Missing Women

Emily Oster*

Harvard University

September 29, 2005

Abstract

In many Asian countries the ratio of male to female population is higher than in the West – as high as 1.07 in China and India, and even higher in Pakistan. A number of authors (most notably Sen, 1992) have suggested that this imbalance reflects excess female mortality and, as a result, have argued that as many as 100 million women are "missing." This paper proposes an explanation for some of the observed over-representation of males: the hepatitis B virus. I present new evidence, consistent with an existing scientific literature, that carriers of the hepatitis B virus have offspring sex ratios around 1.50 boys for each girl. This evidence includes both cross-country analyses and a natural experiment based on recent vaccination campagins. Hepatitis B is common in many Asian countries, especially China, where some 10 to 15% of the population is infected. Using data on viral prevalence by country as well as estimates of the effect of hepatitis on sex ratio, I argue that hepatitis B can account for about 45% of the "missing women": around 75% in China, between 20% and 50% in Egypt, and West Asia, and under 20% in India, Bangladesh, Pakistan and Nepal.

1   Introduction

The ratio of men to women in the Western world is close to unity. Sex ratios at birth are bit higher (around 1.05 boys for each girl), but higher male mortality results in sex ratios around 1.00 by young adulthood and even lower among older adults. Although this pattern is universal in the West, population sex ratios in a number of Asian countries seem to be much higher – 1.07 in China, 1.08 in India and as high as 1.11 in Pakistan. This imbalance has

1

caused a number of authors to argue that there is substantial excess female mortality in Asia (Sen, 1992; Coale, 1991; Klasen, 1994). In general, they argue that neglect of female children and poor conditions for women contribute to the higher sex ratio. These authors have estimated that there are between 60 and 100 million "missing women."

This paper connects the missing women puzzle to a separate literature investigating a connection between the hepatitis B virus and offspring sex ratios. There is evidence that women who are carriers of hepatitis B give birth to a higher ratio of boys to girls than non-carriers. Since many of the countries with missing women also have relatively high hepatitis B carrier prevalence, the naturally occurring higher sex ratio at birth could produce a higher population sex ratio even in the absence of excess female mortality. Indeed, I argue in this paper that after adjusting for differences in sex ratio at birth caused by hepatitis B, the number of missing women drops to 32 million, from the 60 million calculated by Coale(1991) and the 107 million suggested by Sen(1992). There is significant variation among countries in the share of the gender bias that is accounted for: I find that hepatitis B can explain 75% of the missing women in China, but less than 20% in India, Pakistan and Nepal. Other countries lie between these polar cases.

The issue of the missing women has become a flashpoint for development economists, demographers and policy-makers. For many, presumed excess female mortality is only the most severe example of discrimination against women in general in these regions. The missing women are emblematic of what seems to be a much larger problem, and understanding why these populations have such a widely skewed sex ratio seems vital to moving forward on issues of gender equality. The results here suggest that sex ratio bias may be less attributable to differences in tastes and preferences than many have suggested. However, there are at least two important caveats to the results here.

The first is that technological developments in the last 15 years – in particular, ultrasound – have made selection of a child's gender before birth possible in a number of the countries in question here. And, as has been pointed out by many authors, the sex ratios in India and China in particular have correspondingly increased (see, for example, Dreze and Sen, 2003; Hull, 1990; Junhong, 2001). As the parental ability to select the sex of their child becomes less costly, any naturally occurring bias in the sex ratio becomes less important. In

the extreme, if selection was costless, these biases would have no effect. This suggests that, while there may well be, a role for hepatitis B in understanding the existing gender imbalance in the missing women countries, this role will decline in future generations as parental control improves. Indeed, this has already happened to some extent: the increases in sex ratio at birth in China and India in the last 15 years cannot be attributed to hepatitis B. This is particularly true since vaccination for hepatitis B has become virtually universal, but it would be true regardless.

The second, and perhaps more important caveat, is that what is estimated in this paper is the total effect of hepatitis, not the partial effect. The estimates here assume that, in the absence of hepatitis B, the number of excess female deaths would be the same – there would no compensating behavioral response. This is unrealistic. It is clear that there is significant gender discrimination in these countries – the changes described in the above paragraph clearly point to latent demand for male children and, further, even the estimates in this paper suggest about half of the missing women *are* a result of discrimination. If the naturally occurring sex ratio at birth had been lower historically, we would expect some compensating behavioral response.

The correct counterfactual is therefore not the West, but a world in which there is no hepatitis, but a similar level of gender bias. Creating this counterfactual is difficult. In Section 5.2 I attempt to use cross-country variation in hepatitis rates to explore this, but even this requires some significant assumptions. It seems likely that in the absence of hepatitis sex ratios would be somewhat lower, but it is perhaps not clear exactly how much. For this reason, it is important to keep in mind exactly what the estimates represent. Although I will refer to the share of the gender imbalance "explained by" hepatitis, this should be taken to refer to the total effect, not the partial one.

Section 2 of this paper discusses how the number of missing women is calculated. Section 3 then presents a discussion of the origin of the sex ratio bias. The hepatitis B hypothesis would suggest that the sex ratio bias arises at birth; to the extent that the bias arises later, there cannot be a role for hepatitis B. I present evidence that sex ratio at birth is substantially higher in most of the missing women countries than it is in the West. In addition, sex ratio patterns by age vary across these countries. China, for example, exhibits

very high sex ratios at birth and declining sex ratios over childhood while India exhibits moderately higher sex ratios at birth but also sees increasing sex ratios during childhood. The overall high sex ratio at birth suggests a role for hepatitis B in the missing women puzzle, especially in China.[1]

Section 4 presents a variety of evidence pointing to an effect of hepatitis B on sex ratios at birth. Section 4.2 discusses individual-level evidence drawn from the existing literature. In 1972, researchers studying hepatitis B in Greece noted that the sex ratio among 131 children born to women with hepatitis B was 1.77 boys for each girl, versus only 1.13 among 542 children born to women without hepatitis. Five additional studies in different areas also suggested this connection. The evidence from the scientific literature is suggestive, but not necessarily conclusive. As will be discussed in more detail later, the sample sizes are quite small, and there are some issues with data quality. Further, the studies (for the most part) were not designed to test the hepatitis-sex ratio connection and, as a result, the methodologies are not perfectly suited to testing this theory. I therefore devote a significant portion of the paper to new analyses designed to test the hepatitis B-sex ratio connection.

First, Section 4.3 takes advantage of the availability of a vaccination since the late 1970s as a natural experiment. I consider first offspring sex ratios for Alaskan Natives (historically high hepatitis) and non-Natives in Alaska (low hepatitis). Data on births to these groups are compared before and after the vaccination campaign. I find that sex ratios among white Alaskans are unaffected, but sex ratios among Alaskan Natives drop dramatically over the period being considered. In addition, I consider sex ratio among women in Taiwan who were affected by a vaccination program that began in 1984; there is suggestive evidence of a drop in sex ratio at birth.

Section 4.4 presents aggregate evidence on the cross-country relationship between hepatitis B and sex ratio. This section takes advantage of both a categorical distribution of hepatitis B and continuous estimates of hepatitis B prevalence across countries. Both types of evidence suggest that hepatitis B and sex ratio at birth are strongly related, and this finding is robust to restricting to countries within a particular region, countries within the OECD,

---

[1]Recognizing the recent increases in sex ratio at birth in China, I present a number of pieces of historical data and argue that the sex ratio was high in the past, even if not as high as at present.

and to controlling for income levels and regional fixed effects. I find about 40% of the cross country variation in sex ratio at birth can be explained by hepatitis B. In addition, the magnitude of the effect implied by the cross country data is very close to the effect size in the individual-level data. Taken together with the original scientific data, this new evidence is more strongly supportive of the hepatitis-sex ratio connection.

In Section 5 I combine data on hepatitis B prevalence and estimates of the effect of hepatitis on sex ratio at birth to estimate and adjusted number of missing women. The baseline results suggest around 45% of the missing women estimated by Coale(1991) are explained by hepatitis. The hepatitis B-adjusted number of missing women is 32 million – 8 million in China, 19 million in India and smaller numbers elsewhere. In Section 5.2 I attempt to use within-country variation in China to create a more realistic counterfactual for that country. Although the number of missing women under this counterfactual is smaller, the results still suggest around 62% of the missing women are explained (down from 75% in the baseline specification).

First and foremost, the findings in this paper have implications for the puzzle of the missing women. However, the work here is also closely related to work on the economics on the family. In particular, understanding the contribution of biology to gender imbalance necessarily requires understanding the contribution of behavior, and visa versa. Drawing on Becker(1981), a number of papers have developed theories that provide a rational motivation for sex-selective mistreatment, infanticide and abortion, based on future earnings capacity of children, and done empirical work to test those theories (see Rosenzweig and Shultz, 1982; Klasen, 1996; Agnihotri, 1999; Edlund, 1999; Kojima, 2005). For the most part, the empirical work in these papers relies on regressing sex ratios by region on economic factors of interest (maternal education, earnings power, etc). Others (Sen, 1998; CIA World Factbook) have suggested that sex ratio may be used as a direct measure of gender discrimination. Clearly, the validity of the previous regressions and the value of sex ratios as a measure of discrimination may be affected by the contribution of biology to the gender imbalance.

Similarly, good estimates of the effect of behavior on sex ratios may speak to the potential importance of biology. Qian (2005) provides substantially better identified estimates of the effect of economic incentives on sex ratios of surviving children in China. Using

variation across provinces and time in the value of female labor, she shows that increases in the value of female labor increase the share of girls who survive. The Qian (2005) results strongly support the existence of behavioral responses to economic incentives over and above any biological differences. These results are important to keep in mind in the context of this paper, in particular when thinking about appropriate counterfactual sex ratios. However, the results in the two papers are not necessarily at odds; among other things, the hepatitis B hypothesis is about the average sex ratio, and the estimates in Qian (2005) represent marginal changes.

Moving a bit further afield, the results in this paper may also be related to economists' study of marriage and labor market outcomes. In his work on the economics of the family, Becker (1981) highlights the importance of population sex ratios in the formation of polygamous societies, as well as in the determination of dowry and brideprice levels. Building on this, others have argued that high sex ratios (many men in the population) lead to higher rates of marriage among women and lower outside labor market participation (Angrist, 2002; Chiappori, Fortin and Lacroix, 2002). Population sex ratio movements during wars have also been connected to female labor market participation and wages (Goldin, 1991; Acemoglu, Autor and Lyle, 2004). The connection between hepatitis and sex ratio therefore has the potential to provide a cleaner test for the relationship between population sex composition and market outcomes.

2 Calibrating the Missing Women

The population sex ratio in Europe and the United States – number of men divided by number of women – is around one. That is, there are roughly equal numbers of men and women. Sex ratio at birth weakly favors men (about 1.05 men for each woman), but higher male mortality over the lifetime results in equalization of the ratio in the population overall. In Asia, population sex ratios are higher. In China, for example, the population sex ratio in 1990 was around 1.066 men for each woman and was even higher in the Indian sub-continent.[2]

---

[2]Throughout the paper I refer to both population sex ratio and the share of men. Obviously these are isomorphic: sex ratio is simply the number of men in the population divided by the number of women. As a result, both higher sex ratios and higher male shares imply more men in the population.

It was this observation, coupled with evidence about son preference in areas of Asia, that led demographers to first suggest that there were "missing women" in Asia.

"Missing women" are defined as women who once lived but have died prematurely because of artificially high female mortality. The number of missing women in a country is therefore calculated by assuming that the number of men reflects appropriate mortality patterns and using the difference between the actual and expected sex ratio to calculate the level of excess female mortality. Denote the expected sex ratio based on the number of men as SR_e, the actual sex ratio observed as SR_a and the number of women observed as N_w. Then the missing women can be calculated as below:

$$Missing = \left(\frac{SR_a}{SR_e} - 1\right) N_w \tag{1}$$

Obviously, the larger the difference between the actual and the expected sex ratio, the greater the share of women who are missing. Conversely, if $SR_a = SR_e$ then there are no missing women.

From the perspective of this paper, the origin of the sex ratio bias is crucial. Parental hepatitis B infection is posited to affect offspring sex ratio at birth, which will then play into later sex ratio of the population. In particular, overall sex ratio of a given cohort in a population is a function of their sex ratio at birth and their mortality patterns by gender. Denoting the actual sex ratio in cohort c as $SR_{a,c}$, sex ratio at birth as $SR_{b,c}$, male mortality for the cohort to this point as M_m and female mortality M_f, I can write:

$$SR_{a,c} = SR_{b,c} \frac{1 - M_m}{1 - M_f} \tag{2}$$

It is easy to see, then, that even if two populations have identical mortality by gender, if their sex ratios at birth differ then the sex ratio later in life will differ also. A higher sex ratio at birth will produce a higher population sex ratio, as will more imbalanced mortality patterns.

In Section 4, I present evidence suggesting that the offspring sex ratio is higher for hepatitis B carriers than non-carriers. Assuming that carriers have a higher sex ratio among offspring, then the prevalence of hepatitis in the population will influence the sex ratio at

birth. More specifically, the sex ratio at birth will be the average of the sex ratio of hepatitis B carriers (SR_{HBV}) and non-carriers (SR_N), weighted by their population shares:

$$SR_{b,c} = (SR_{HBV})(Prev_{HBV}) + SR_N(1 - Prev_{HBV}) \qquad (3)$$

The final step in connecting hepatitis B and the missing women is to calibrate the effect of hepatitis on sex ratio – to calculate SR_{HBV} from the above equation. For the purposes of calibration, I use estimates of the effect of hepatitis B on sex ratio at birth that are calculated from the individual level data. I create a dataset based on the individual level studies, assuming each child is an observation. Child gender is regressed on parental hepatitis B carrier status, with a fixed effect for each study. This is discussed in more detail in Section 4.5.

As a robustness check, I also calculate the SR_{HBV} suggested by aggregate population level data. Consider two populations: Population A with a high hepatitis rate H_A and a high sex ratio SR_A and population B with a lower hepatitis rate H_B and a lower sex ratio SR_B. Assuming that the non-carrier sex ratio is the same in the two areas, solving for SR_{HBV} yields the following expression:

$$SR_{HBV} = \frac{SR_A(1 - H_B) - SR_B(1 - H_A)}{H_A - H_B} \qquad (4)$$

The discussion above provides a framework for the rest of this paper. I now move to discussing the origin of the sex ratio bias in the missing women countries, and then to the possible role of hepatitis.

3   The Origin of the Sex Ratio Bias

This paper argues that hepatitis B may account for some of the missing women. This argument, however, relies crucially on the sex ratio bias arising at birth.[3] It is worth noting here that what may seem like relatively small differences in sex ratios at birth can result in a

---

[3]It is worth noting that men are more likely to be Hepatitis B carriers and therefore may have higher hepatitis B-related mortality from liver disease. However, this effect will already be accounted for in the analysis, since the life tables used for the calculations adjust for differences in female and male life expectancy.

large sex ratio bias in the overall population, similar to what is seen in Asia. For example, moving from a sex ratio at birth of 1.05 to 1.10 and keeping mortality patterns constant moves the overall population sex ratio from 1.00 to 1.05, which is a large majority of the U.S./China difference.

To get some sense of sex ratios over the life cycle in the missing women countries, Figure 1 shows the sex ratio by age category in each country. This figure is created by calculating the sex ratio by age group from data in the Demographic Yearbook Historical Supplement (United Nations, 1997).[4] The data used are from censuses or birth registration data taken between 1970 and 1990. It is worth noting, of course, that this methodology combines cohort and age effects and if there are changes in population demographics over time, this will be problematic; below I address this issue below for India and China by following individual age cohorts over time.

The pattern in sex ratios over the life cycle appears to vary across countries. In China, for example, sex ratio is highest at birth and then generally decreases, although is slightly higher among those 35-59 than among those aged 15-34. The pattern in India is somewhat different: sex ratio is slightly higher than "normal" at birth and increases in childhood and young adulthood before falling again for the very old. Data for Pakistan and Bangladesh suggest a slightly different pattern, with above average sex ratios at birth, increases in childhood and large increases in old age. The pattern in West Asia (Turkey and Syria) and Egypt is more typical of developed countries, but with slightly higher sex ratios at birth than those seen in the West.

Figure 1 generally suggests that the origin of the sex ratio bias is not the same in every country (although, as mentioned, this figure confuses age and cohort effects). Although most of these countries seem to have higher sex ratios at birth than the average in the West, many of them also have non-Western mortality patterns, even controlling for their income level. The increase in sex ratio in childhood – particularly in India, Pakistan and Bangladesh – suggests excess childhood mortality for girls. In contrast, in China, the data indicate high

---

[4]For India and Nepal, sex ratio at birth is not available from the Demographic Yearbook, so household survey data are used to construct a measure of the sex ratio at birth for these countries based on reported sex ratio among ever born children. The reliability of these data is checked using data from Pakistan, for which both household survey data and Demographic Yearbook data are available, and the results suggest that the household survey data provide a reasonable estimate.

sex ratios at birth but mortality patterns that resemble the West's.

In order to avoid the confluence of age and cohort effects, Figure 2 follows individual cohorts in China and India rather than considering age specific sex ratios (I have chosen these countries because they represent by far the most missing women and have different sex ratio patterns by age). For example, the graph includes sex ratio by age for the 1982 birth cohort in China – their sex ratio at birth and then in subsequent years. In general, these data are from less-than-complete censuses (in China, for example, these are data from the yearly 1% censuses), but it should give a good sense of the sex ratio patterns over the early part of the life cycle. Figure 2 mirrors Figure 1: The sex ratio bias in China is clearly evident at birth, and only declines through childhood, while the sex ratio in India is low at very young ages, and increases through childhood.

Figures 1 and 2 generally demonstrate that there appears to be some sex ratio bias at birth, at least in a subset of the missing women countries. In particular, in most of these countries the average sex ratio at birth is above 1.059, which is the fixed sex ratio that Coale (1991) uses to calculate the number of missing women. As noted in Section 2, increases in sex ratio at birth will lead directly to increases in the sex ratio later in life. When re-estimating the missing women, I assume exactly the same mortality patterns as in Coale (1991), and use hepatitis B-adjusted sex ratios at birth to explore the effect on overall sex ratios.

In general, the results in this paper suggest that the number of missing women in China in particular has been overestimated. Given this, it is worth briefly exploring whether the sex ratio bias at birth in China is naturally occurring. It has been argued that the sex ratio bias in China has been rising over time and that imbalanced sex ratios are due largely to sex-selective abortion (Hull, 1990; Yi et al, 1993; Junhong, 2001). While measurement is extremely difficult (for one attempt on a small scale, see Junhong, 2001), there can be no doubt that there has been some use of sex-selective abortion technology in China. For example, the sex ratio at birth in the 2000 Census is close to 118 boys for 100 girls, much higher than that reported in 1989 or 1982, and even in 1995. The interaction between the one-child policy and the increased availability of ultrasound technology has had some effect on these ratios. For example, authors have noted higher sex ratios among later children, although this does not seem to be strongly true earlier in the century (Coale and Banister, 1994).

However, the fact that sex ratio at birth has increased over the last 15 years does not necessarily imply that sex ratios prior to this time period were normal. The missing women estimates presented in Sen(1992) and Coale(1991) rely on population sex ratios estimated between 1981 and 1991. In general, most or all of the cohorts alive in those years would not have been affected by either availability of sex-selective abortion, or the one-child policy. As a result, what is relevant for comparing to these estimates is the sex ratio at birth before the increased availability of sex-selective technology. Here, I present a number of pieces of evidence suggesting that sex ratio at birth has been high in China historically and, more generally, evidence that suggests the bias at birth prior to the latest censuses is organic.[5]

In an early study of the effect of geography on offspring sex ratio in humans, Chambliss (1949) notes the generally high sex ratio at birth in China. In five regions of China, Chen (1947) reports sex ratios at birth that range from unity to 123 men per 100 women. Most are substantially higher than normal, and the average is 108 men per 100 women. Chiao (1934) reports a ratio of 110.4 boys for 100 girls in his study of 12,000 farm families in 11 provinces between 1929 and 1931. Chiao (1934) additionally reports the results of three sources of hospital registration (average sex ratio of 111.6 boys for 100 girls), three registration surveys (average sex ratio of 114.7) and three general surveys (average sex ratio of 115.2). It is notable that even the hospital registration reports a very high sex ratio in this time period in the late 1920s.

These papers, however, generally report the results of particular censuses and individual studies that focus on one time period. In contrast, Coale and Banister (1994) do an extensive study of the path of sex ratios at birth over five decades in China. Using two retrospective fertility surveys from the 1980s, they report imputed sex ratios at birth for cohorts born between 1936 and 1989, and find high sex ratios at birth from the earliest

---

[5]The first thing to note is that sex-selective abortion technology (ultrasounds) were not widely available in China prior to the 1980s (Junhong, 2001) and they dispersed over the 1980s and 1990s. The preceding analysis uses sex ratio at the 1982 and 1989 birth cohorts; this first cohort is unlikely to be heavily influenced by this technology, and yet the sex ratio is abnormally high (close to 1.09).

cohorts.[6]

An alternative way to explore the importance of underreporting and infanticide in China is to look at the offspring sex ratios of native Chinese women in the United States. There is good evidence that hepatitis B rates among Chinese immigrants to the West are similar to those in China (Kent, 2000; Gjerdingen and Lor, 1997; Hayes et al, 1998), and it is difficult to believe that there is infanticide among these groups in the United States.[7] This suggests that high sex ratios among children of these immigrants would support the hypothesis of naturally occurring high sex ratios at birth.

I explore two measures of sex ratios at birth for children of Chinese immigrant mothers: vital statistics and the U.S. Census Individual Public Use Microdata Sample (IPUMS). The Center for Disease Control Vital Statistics Natality Detail Files report on all births in the U.S. from 1985 onwards. Information on sex of the child, race and origin of the mother and father, as well as basic demographic information is included. The main drawback from the perspective of this exercise is the extremely limited data about national origin and race. It is possible, however, to estimate sex ratios among children born to mothers whose race is "Chinese" and who were not born in the United States. There are two important caveats: first, the racial coding of "Chinese" includes not only those who identify themselves as Chinese, but also those from Taiwan, Macao and other places, which may have very different (and lower) hepatitis B rates. Second, since birthplace outside the United States is identified only as "rest of the world," it is hard to know whether women were actually born in China. Both of these factors will bias the sex ratio downwards relative to that for Chinese women immigrants from mainland China, but the sex ratio at birth is still substantially higher than normal, at 1.082. This is strongly significantly different from the sex ratio for whites.

---

[6]Table 1 reports some of the data from their paper, showing sex ratios at birth starting with the 1936-1940 cohort and going through the 1985-1989 cohort. The sex ratio is high in all cohorts. Although it declines a bit in the 1960s, the overall average is 1.095. It is worth noting that, although retrospective data can be problematic, Coale and Banister (1994) provide evidence suggesting that the retrospectively reported sex ratios at birth line up well with actual sex ratios from later censuses, suggesting that the data are reliable.

[7]Concerns about selective migration should be somewhat alleviated both through information on high hepatitis B rates among immigrants and by the fact that there seems to be little evidence that hepatitis infection is related to socioeconomic status. Lee et al (2002) find a weak positive association between hepatitis and income in Korea; Wang et al (2002) echo these findings in Taiwan; Rahman et al (1997) find no relationship in Bangladesh; Amini et al (1993) find a negative relationship in Iran and She et al(1988) find no relationship among barbers in China. Since hepatitis B carrier state does not generally cause active illness, there is no good reason to believe carriers would be less likely to migrate than others.

12

In contrast to the vital statistics, the IPUMS data (which give detailed information on 1% or 5% of U.S. households) provides very detailed information about country of origin; the downside is a smaller sample size and the fact that it is necessary to use data on sex ratios among young children rather than at birth. With limited infant mortality the sex ratio among young children should mirror the sex ratio at birth, so this may not be a concern, particularly in later time periods. Panel A of Table 2 shows sex ratios for U.S. born children under 6 with Chinese mothers from the IPUMS. The sex ratio overall is quite high – 1.105. In the last three censuses, it is even higher, at 1.115. In all periods except 1940-1970 the sex ratio is very high. As a comparison, Panel B presents the sex ratio of births to immigrant mothers from low-hepatitis B countries, and the sex ratio of births to native whites. Both are normal – around 1.04 – and significantly different from the rate for Chinese mothers. The comparison to other immigrant mothers is particularly noteworthy, as it suggests that the high sex ratio among Chinese mothers is not driven by their immigrant status.[8]

4   Hepatitis B and Offspring Sex Ratio

This section discusses the hepatitis B virus and the effect of the virus on offspring sex ratio. The first subsection provides background on the virus; I then discuss the evidence for an effect of hepatitis B on offspring sex ratio. The final subsection calibrates the size of the effect of hepatitis B on sex ratio and uses data on hepatitis B prevalence to estimate the share of the bias in sex ratio at birth that is explained by hepatitis rates.

4.1   Background on Hepatitis B

Hepatitis B virus (HBV) is a viral infection of the liver that is spread through exchange of blood or bodily fluids. The virus can be spread through sexual contact, needle exchange and from mother to child, as well as through more casual contact (sharing saliva, etc). There is an important distinction between ever being infected with HBV and being a carrier. Many infected individuals clear the virus from their system in a relatively short period of time;

---

[8]I elaborate on this in Section 4.4, which demonstrates that hepatitis B prevalence in the home country strongly predicts sex ratios at birth for immigrant women in the IPUMS from a wide variety of countries.

however, for some infected individuals – even those who do not show symptoms – the virus does not clear and they go on to become chronic carriers. Chronic viral infection leads to serious liver problems for around 25% of carriers. The chance of becoming a carrier of the virus conditional on infection decreases with age of infection – it is close to 90% for infants, but under 10% for those infected as adults (Department of Vaccines, 2001). Since the late 1970s a vaccine for HBV has been available. It has been widely used in a number of places – Alaska, China, Taiwan, Singapore and some countries in Southern Europe – although it is rarely used in countries where HBV prevalence is not high.

Cross-country HBV rates vary significantly. High hepatitis regions include East Asia, Sub-Saharan Africa, parts of the former USSR, parts of Eastern Europe, northern South America, Alaska and northern Canada. From the perspective of this paper, it may be important to understand why HBV carrier rates vary across space – for example, how important is the frequency of risky sexual behavior in carrier rates?

In Appendix A I present a simplified model of the transition path of the virus across generations and use it to inform both the question of cross-space and cross-time variation in carrier rates. As is mentioned above, the variation in the chance of becoming a carrier (given infection) over the life cycle is crucial. Infants who are infected by their mothers have a 90-100% chance of becoming a viral carrier; children who are infected by siblings or playmates have around a 30% chance of becoming a carrier; adults who are infected have only a 5-10% chance of becoming a carrier. Given these facts, it is shown in Appendix A (and should be intuitively clear) that differences in the share of infants and children infected are crucial to differences in carrier rates.

The necessary follow-up question is why infection rates among infants and children vary across space. The answer here is less clear, but it may have to do with differences in viral or human genotype. Hepatitis B has eight known genotypes (A-H) and there is some circumstantial evidence that their distribution is correlated with HBV prevalence (Miyakawa and Mizokami, 2003) and that the genotypes differ in their transmission patterns (Duong et al, 2004; Kobayashi et al, 2003). Human genetic differences in susceptibility are also a possible explanation for the variation in epidemicity. Geneticists have identified a number of

polymorphisms[9] that appear to affect susceptibility to HBV (for a review, see Andrade and Andrade, 2004). One implication of this, and of the discussion in Appendix A, is that choices about sexual behavior are unlikely to play a major role in determining carrier rate.[10]

Related to the issue of cross-space variation is the issue of cross-time variation. Implicity, when adjusting the number of missing women in Section 5, I assume that the sex ratio at birth in all currently living cohorts was the same. If carrier rates have not been stable over time, this may be unrealistic. The same model of viral transmission that is used in Appendix A can be used to infer whether it is likely – given the transmission path and the age of the virus – that current infection rates are a steady state. In general, the results there indicate that the system would move relatively quickly into steady state, generally in less than a tenth of the time since the virus was introduced, implying that current infection rates are likely to represent long-established stability.

4.2 Individual-Level Evidence

The relationship between HBV carrier state and offspring sex ratio was first noticed in four studies on families of carriers and non-carriers (Hesser, Economidou and Blumberg, 1975; Drew, London, Blumberg and Serjeanston, 1982; Drew, Blumberg and Robert-Lamblin, 1986; Chahnazarian, Blumberg and London, 1988). The methodology in these studies is relatively simple. In each case, married women with children were identified in villages with known high HBV rates. A sample of women and (in all cases except Greenland) their husbands were tested for HBV carriage and were given a detailed survey about their reproductive history. Subsequently, two additional studies (Cazal, Lemiare and Robinet-Levy, 1976; Livadas et al, 1979) specifically sought to test the results using the same methodology, and came to the same conclusions. All six studies tabulate the gender of all children of carrier and non-carrier parents and compare the sex ratio across the two groups.[11] It is worth noting that there is no

---

[9]Polymorphisms are mutations in the human genetic code that appear in at least 1% of the human population. Human blood groups are an example.

[10]In fact, HBV prevalence across countries seems to be largely uncorrelated with the prevalence of other sexually transmitted infections; for example, the correlation with herpes prevalence in a subset of the countries in the sample is around -0.06.

[11]There is one further study of this issue that employed a slightly different procedure (Mazzur and Watson, 1974). In the British Solomon Islands families were tested for HBV and the sex ratio among carriers and siblings of carriers was explored. Assuming that the transmission of the virus to the carriers was parent-to-child, then

a priori reason to expect carrier and non-carrier families to be different from each other, absent HBV. In nearly all cases adults would not know about their HBV status and would have no symptoms during childbearing years.

The estimated offspring sex ratios for carrier and non-carrier parents in each of the six studies cited above are presented in Table 3. Overall, families with an HBV carrier parent had 355 male children and 248 female children (sex ratio of 1.44), while families without an HBV carrier parent had 2083 male children and 2043 female children (sex ratio of 1.02). The difference is statistically significant overall and in each individual study.

Although individual-level data is the most straightforward way to test the hepatitis-sex ratio connection, the existing evidence has significant problems. First, the sample sizes are quite small – in most cases not more than a few hundred children total – and the studies took place in only a limited set of locations.

Further, the first four studies in Table 3 were not specifically designed to test this theory and, for this reason, the data are not ideal for the purpose. As discussed more extensively in Chahnazarian et al (1988), the data quality varies significantly across studies, and is particularly problematic in Greenland and Kar Kar Island. They also worry that the high sex ratio among non-carriers in the Greece 1 study and the low sex ratio in the Kar Kar Island study are results of sample selection or systematic under-reporting. As I discuss more extensively in Section 4.5, this problem is somewhat (although not completely) mitigated by the inclusion of study-specific fixed effects in the regressions. Despite this, low data quality is problematic in general, and may point to more serious issues. Some comfort is provided, perhaps, by the Greece 2 and France studies, both of which were designed specifically to test the hepatitis-sex ratio connection, and for which the methodologies are more suited.

A third, and perhaps more obvious, concern is that the mechanism by which hepatitis affects the offspring sex ratio is not known. There are several hypotheses in the literature; for example, a number of papers have suggested that female fetuses of HBV carrier parents are much more likely to spontaneously abort early in the pregnancy, resulting in more male live

---

the families with siblings who are carriers will be offspring of HBV-positive individuals. In this study, the sex ratio among siblings of carriers was much higher than among siblings of non-carriers (1.61 versus 0.97). I have not included this study because the parents were not tested directly, but it is comforting that the magnitude of the effect is similar.

births (Drew et al, 1978; Livadas et al, 1979), but this has not been medically confirmed. An alternative theory suggests that carrier women are less likely to conceive daughters. Without a direct mechanism, the weaknesses of the individual-level data become more problematic and further confirmation seems necessary.

To summarize, while the existing scientific evidence is clearly suggestive of a hepatitis-sex ratio connection, it is not conclusive. In the next two sections of the paper I use new data and new methodologies to test this connection. The paper will not endeavor to prove the biological mechanism. Instead, I rely on population level data and methodologies that will be familiar to economists to test the strength of the hepatitis B-sex ratio relationship. In the next section, I use a natural experiment suggested by the hepatitis B vaccination campaign to estimate the effects; section 4.4 presents further cross country evidence.

4.3   Time Series Evidence

A vaccine for HBV became available very shortly after the virus was discovered in the late 1970s. To the extent that vaccinated women are now of childbearing age, this provides a natural experiment to explore the effect of HBV on sex ratio. In general, universal vaccination programs (in Singapore, Italy, Spain, etc) did not start until the mid-1980s or early 1990s. However, there are two areas that allow me to consider variation over time: Alaska and Taiwan.

The Native Alaskan population has historically been the only population in the U.S. with high hepatitis rates – as high as 16% to 20% in some villages, while non-Native Alaskans have very low HBV rates. Further, HBV prevalence varied considerably across Alaska (McMahon et al, 1993), with particulary high prevalence in West and Central areas, and much lower in the Far North and East. The U.S. began vaccination of the entire Native population in 1982 and the vaccination campaign included not only infants but adolescents and adults (Harpaz et al, 2000).

Unlike in vaccination campaigns that focus only on infants, the overall coverage of this campaign meant that the incidence in the population overall fell (effectively to zero) immediately. The share of carriers (prevalence) among women of childbearing age decreased gradually, since anyone who was infected already would remain infected even after the

vaccination campaign. However, by the end of the 1990s nearly all women who were giving birth would have been vaccinated and therefore would not be carriers. In addition, over the period from the early 1980s through the late 1990s the number of infected women giving birth would have decreased gradually, as carriers were replaced by vaccinated individuals in the population of women of childbearing age. It may therefore be possible to see changes in offspring sex ratio over time, even considering only the last 20 years.

Data on the number and gender of births are available across regions in Alaska from 1980 through 2000. I compare the changes in sex ratio at birth across three groups: high-native, high-HBV regions (Yukon-Koyukuk, Aleutians East Borough, Lake and Penninsula, Bristol Bay), high-native, low-HBV regions (Prince of Wales, Skagway-Hoonah-Angoon, Wrangall Peterson) and a low-native region (Anchorage). The effect of the vaccination program on these three groups can be seen in Figure 3.

While sex ratios are relatively constant in the high-native, low-HBV and low-native regions, they decrease dramatically in the high-native, high-HBV region. The sample sizes are small, but it is possible to reject equality between the normal sex ratio at birth and the ratio observed in the high-native, high-HBV regions in the earliest time periods.

It is also possible to use the IPUMS to explore the changes in sex ratio for young children among the two groups over this time period. I consider the sex ratio among very young children born to whites and Natives in Alaska from the 1980 through 2000 censuses. These data span the time period preceding and following the vaccination campaign. I expect a similar pattern to that seen in the vital statistics: decreases in sex ratio among natives between 1980 and 1990 and then between 1990 and 2000. More specifically, I estimate the following equation:

$$male_{i,t} = \alpha + \beta_1 nat_{i,t} + \beta_2 d1980_{i,t} + \beta_3 d1990_{i,t} + \beta_4 nat_{i,t} * d1980_{i,t} + \beta_5 nat_{i,t} * d1990_{i,t} + \mathbf{X\Psi}$$

where $male_{i,t}$ is a dummy for whether child i observed in census year t is a boy and $nat_{i,t}$ is a dummy for whether that child is an Alaskan Native. In addition, there are two year dummies – a dummy for census year 1980 ($d1980_{i,t}$) and a dummy for census year 1990 ($d1990_{i,t}$). The omitted category is the 2000 census. The coefficients of interest are β_4 and β_5, which measure the additional positive effect of being a native on the probability of being a boy in the 1980

and 1990 censuses. The theory suggests that the effect of being a native Alaskan on the probability of being male should be highest in 1980, and then decreasing to 1990 and 2000 (so both β_4 and β_5 should be positive, and β_4 should be greater than β_5). **X** is a vector of controls, including a full set of maternal and child age dummies, and dummies for child year of birth.

Column 1 of Table 4 reports the results of this regression. The overall effect of native is zero, but the interactions are both positive and significant. Sex ratio among natives is higher than that among non-natives in the 1990 census, and higher still in the 1980 census. In Column 2, I limit to Eskimos, the ethnic group which is generally thought to have the highest HBV rate (McMahon et al, 1993). Although sufficiently detailed racial identifiers are not available until the 1990 census it is nevertheless possible to get a sense of the pattern using 1990 and 2000. In this case, the dummy on Native is positive, as is the interaction: this ethnic group has a higher offspring sex ratio during both 1990 and 2000, but it is even higher in 1990 (when the vaccination program would have covered fewer people) than later.

There is some similarly suggestive evidence on HBV and sex ratio from Taiwan. Taiwan was historically among the highest HBV regions in the world (with a carrier rate of 15% to 20%) and began universal vaccination of at-risk newborns in July of 1984, and then of all newborns in 1986 (Lin et al, 2003). In addition, some catch-up vaccination was done among school-age children and adults in the late 1980s (Ni et al, 2001). The extensive vaccination program in such a high risk country suggests another opportunity to explore the effect of vaccination on sex ratios. Data on the gender of births by maternal age group (15-19, 20-24, etc) are available for all births in Taiwan (Republic of China, 1974-2002). Since women who are vaccinated in 1984 would be 15 by 1999, there is potential scope for exploring the effect of the vaccination program on births to young mothers late in the 1990s.

In theory, one could use the discontinuity associated with vaccination introduction in 1984 to analyze births to vaccinated and non-vaccinated mothers. This is not possible for two reasons. First, the catch-up vaccination among school-age children means that there are serious limits to the discontinuity. Secondly, and more importantly, there will always be "bleeding" of the vaccination program among children of similar ages. Given that a large portion of infections in any given cohort will be a result of horizontal (child-to-child) contact, vaccination among children born in July 1984 will not only lead to lower infection rates

among those children, but also among children born in the months (and potentially years) before them.

This issue is not insurmountable, however. There has been extensive work on the changes in infection rate among vaccinated birth cohorts and unvaccinated cohorts close in age. Using data from a number of papers (Hsu et al, 1999; Ni et al, 2001; Lin et al, 2003) I construct carrier rates by birth cohort. These data indicate that the decrease in carrier rate begins in the 1979 birth cohort, although this decrease accelerates after 1984, as expected. It is worth noting that the bleeding of the program into earlier birth cohorts is potentially advantageous as it implies that older women (for whom there are more birth data) are also affected.

Using the information on carrier rates by birth cohorts and births by maternal age I construct a panel dataset on carrier rates and sex ratio at birth. Data from 1979-2002 are used, and crossed with births to women 15-19 years old, 20-24, 25-29 and 30-34. The final cohort is not affected by the vaccination program, while all others are affected to some extent. The only wrinkle is that, because births are available only by five-year age group for women, average carrier rates for the entire group are constructed by averaging over the birth years included in each cohort.

Before moving on to a discussion of the results, it is important to note a second caveat to this analysis: the vaccination program coincides quite closely with an increased availability of fetal sex-determination technology, and a probable increase in sex-selective abortion. Given a perfect sex selection technology, it is not clear that the vaccination program would have any effect – if women are able to perfectly choose the gender of their child, then decreasing the naturally occurring sex ratio in the womb may not have any effect on the realized sex ratio at birth. However, there is still potential for the program to have an effect if there is variation across the population in the degree of son-preference or in the willingness to abort. In particular, if a share α of the population is willing to abort (and this propensity is at least partially orthogonal to expected carrier status), then there is still scope for decreases in the other $(1 - \alpha)$ share. It is worth noting, however, that the magnitude of the effect of hepatitis on sex ratio at birth is likely to be smaller in this analysis than elsewhere since there is almost surely some countervailing behavior.

Table 5 reports the results of a regression of the percent of births that are male by cohort-year on hepatitis B prevalence. In both columns cohort fixed effects are included; in Column 1 I include linear controls for year and Column 2 includes year fixed effects. In both columns there is a significant effect of HBV prevalence on the share of children that are male; this effect is significant at the 1% level in Column 1, and at the 6% level in Column 2. In addition, as predicted, the effect is smaller than would be implied by the data on individuals presented in Table 3. Here, the implied sex ratio at birth among HBV carriers is around 1.25 boys for each girl; as is discussed in more detail later, both the individual level data and the cross country data imply an estimate of around 1.5 boys for each girl.

Obviously it is still quite early to draw conclusions from the Taiwan experience and more data on future years will be necessary. Nevertheless, the existing data are suggestive and consistent with an effect of HBV on sex ratio at birth.

4.4 Cross Country Evidence

The individual level and time series evidence points to an effect of HBV on offspring sex ratio. If HBV has such an effect, it should be possible to see a cross country relationship between HBV prevalence and the share of boys born.

The World Health Organization (WHO) categorizes countries into high (8% to 20% carriers), medium (2% to 7%) or low (<2%) HBV prevalence. Appendix B lists the countries in each group.[12] These data on hepatitis are combined with data on the share of boys born from the United Nations Demographic Yearbook Historical Supplement (United Nations, 1997). This covers the late 1960s through present.[13]

Figure 4 shows the share of males at birth by WHO category. The leftmost set of columns includes all countries. The share of births that are male in high HBV countries is 51.8%, versus 51.3% in countries with medium prevalence and 51.2% in countries with low

---

[12]Although these data are useful in the sense that it covers all countries, there are drawbacks. The methodology for assigning countries to groups is not made obvious. In addition, in some cases, no studies have been done on hepatitis B in a given country and yet the WHO assigns the country to a group based on its geographic location.

[13]For South Korea and China the data are truncated at 1990, due to clear evidence of increases in sex ratio at birth after that time, likely due to sex-selective abortion. If all data are truncated at 1990 the results look very similar – i.e. South Korea and China seem to be the only places evidencing large increase in sex ratio over time.

prevalence. The middle set of columns presents the world sex ratio excluding China to ensure that the results are not driven by this one high-population outlier and the results are still very strong. The third set of columns limits the analysis to Europe, which has substantial variation in HBV rate but less racial or income level heterogeneity. The share boys-HBV relationship holds strongly even in this limited sample: share of births that are male in high HBV countries is around 51.7%, versus 51.5% for medium and 51.3% for low HBV countries.[14]

The HBV-sex ratio relationship in the WHO data is suggestive but, as discussed, the data may not be very precise. In addition, this type of evidence is limited by the fact that HBV prevalence may vary quite widely even within a category. I therefore move now to consider more continuous measures of HBV prevalence. Although no centralized database with cross-country HBV rates is available, I have created such a database by aggregating published studies that report HBV carrier rates. To this end, HBV estimates for each country were constructed by searching "HBsAg *country name*" in PubMed and extracting all published studies that reported a prevalence rate and sample size and did not include high-risk groups.[15] Overall prevalence rate for each country is a weighted average of the prevalence rates for all studies. A list of studies used is available on the author's web page.

Although data were available for 63 countries, in many cases the total number of individuals tested was very small. To avoid attenuation bias from measurement error as much as possible, the data were limited to countries in which at least 2,500 people were tested. This results in a sample of 38 countries. Figure 5 shows a graph of share male births on hepatitis rate for these countries. The graph is clearly upward sloping, even if China is excluded. In fact, China is not a particular outlier and lies quite close to the regression line. The major outliers are Iran and Mexico, for which the sex ratios at birth are quite low. The data from both countries are known to be incomplete, which may explain this somewhat (I discuss completeness more below). In addition, in Mexico, gender of infants that die within the first few days are not recorded. Since these are much more likely to be boys (due to higher male neonatal mortality) this may explain the low recorded sex ratio. In Figure 6, the data are limited to OECD countries to avoid the suggestion that income and development level

---

[14]All differences discussed here are significant at less than the 1% level in a test of equality of proportions.
[15]HBsAg is the HBV surface antigen that is present in HBV carriers.

differences may be driving the results. The relationship is of a similar magnitude (although slightly weaker statistically) if the data are limited to European countries.

Table 6 reports estimates of the effect of HBV on the share of boys born using these continuous prevalence data. Column 1 reports the overall relationship between HBV and share boys born without any controls. The coefficient is 0.0009, implying that a 1 percentage point increase in HBV prevalence would increase the share of boys born by around 1 tenth of a percentage point. It is worth noting that this is quite a large effect. It means, for example, that increasing the prevalence of HBV from 0% to 10% implies an increase in share of boys born from around 51% to 52% – in the language of sex ratios, this is a move from 1.04 boys for each girl to 1.08 boys for each girl. The coefficient is strongly statistically significant. In addition, HBV alone appears to explain close to 40% of the cross country variation in sex ratios; if Mexico and Iran are excluded the r-squared is 60%.

In the remaining columns the robustness of this result is explored. Column 2 adds a number of controls for socioeconomic status and fertility – these include log GDP per capita, infant mortality, life expectancy and fertility rate.[16] The coefficient is increased very slightly, and more precisely estimated. This is encouraging; in addition to showing robustness to these particular controls, the fact that the coefficient is largely unchanged suggests that socioeconomic status in general is unlikely to seriously affect this relationship.

Column 3 limits the regression to the OECD only; the coefficient is slightly larger, and still strongly statistically significant. In Column 4 I return to using all countries but include regional dummies. The coefficient is unchanged. Finally, in Column 5 I limit the data to country-years in which the United Nations reports a complete census (more than 90% of births recorded). It is not obvious why this is necessary; if a random sample of births are recorded then this should give an unbiased estimate of the share of boys born. However, it is worth considering the robustness of the finding to a restriction of the data. As is clear from Column 5 the results are, if anything, stronger – the estimate almost exactly the same size and significantly more precise.[17]

---

[16]This result is also robust to other measures of socioeconomic status, such as male and female literacy rates.

[17]The analysis is very similar when historical sex ratios – from the first half of the century or earlier – are used rather than present day ratios. This is encouraging, as it suggests that the patterns are long-lived, which is consistent with an explanation relying on an old disease.

Despite the results above, there remains the possibility that cultural differences across countries drive differences in sex ratio and HBV prevalence. In order to somewhat alleviate this concern it is possible replicate the results using data on sex ratios among children of immigrants to the United States. This analysis also allows me to consider additional controls in a way that was not possible in the overall cross country analysis. Using the data from the IPUMS (discussed in more detail Section 3) I link HBV rate in maternal country of birth to the chance that a child is male.[18] In this way it is possible to create a cross-country dataset on sex ratio that is at least not contaminated by cross-country differences in the circumstances of pregnancy and birth.

One possible problem with using data from the IPUMS is the possibility that immigrants are selected on, among other things, low hepatitis rates. To the extent that it is available, the data generally do not support this. In particular, there is good evidence that immigrants from Asia to North America have very high HBV rates.[19] In the case of Asian immigrants, these studies cite HBV rates as high as 20%, which is actually higher than what is typically seen in the home country. Further, as discussed in Section 3 there is little evidence that HBV infection and socioeconomic status are negatively correlated, so the concern that only the rich emigrate to the U.S. and they have lower disease burden probably does not hold in the case of HBV.

The relationship between hepatitis and sex ratio holds in the IPUMS. Figure 7 replicates Figure 5 using the IPUMS data from 1990 and 2000 rather than the actual country data on sex ratios (the sample is somewhat truncated due to data imitations – there are not sufficiently large numbers of immigrants in many countries). The simple relationship between HBV and sex ratio is positive, and significant at the 1% level.

In addition to the obvious advantage of eliminating cross country variation in pregnancy and birth conditions, using the IPUMS allows me to adjust for other

---

[18]This includes all children under 6. As long as I assume that childhood mortality in the U.S. is not different for immigrants from high-hepatitis countries and those from low-hepatitis countries, then using sex ratio among young children rather than at birth will not be a problem. Although using vital statistics would provide sex ratios at birth, the information about country of origin is extremely limited and would make this type of analysis impossible.

[19]See, for example: Kent(2000) on Chinese immigrants to Canada, Gjerdingen and Lor(1997) on Hmong patients in Minnesota, Hayes et al(1998) on pediatric refugees to Maine, Adair and Nwaneri(1999) on African immigrants to Minneapolis and Patel and Voigt(2002) on Vietnamese immigrants to the midwest.

individual-level factors that have been argued to affect the sex ratio, including female social status, marital status and age. To do this, I extend the sample back through the 20th century, and control for year of birth in order to allow more countries to be used in the analysis. In Table 7 I present estimates from a regression using the IPUMS sample, regressing whether or not a child is a boy on HBV prevalence rate in the maternal county of origin and a number of controls. In Column 1 I include only the hepatitis rate measure and dummies for year of birth. The coefficient on prevalence is highly significant. In addition, the magnitude of the coefficient on prevalence is quite close to the coefficient estimated from the actual cross country data.

A number of researchers have noted that maternal age (or birth order) appears to affect sex ratio, with older mothers giving birth to more boys (Lazarus, 2002). If the age distribution at maternity across countries is correlated with HBV prevalence, estimates could be biased. Column 2 of Table 7 includes dummies for maternal age in the regression, with virtually no change in coefficient size – if anything, the effect becomes more pronounced.

The Trivers-Willard hypothesis (Trivers and Willard, 1973) suggests that sex ratio of offspring may be influenced by female status; in particular, females of higher status have more male children. Although some studies have supported this, others have not (for a summary, see Lazarus, 2002). It is worth noting that the hypothesis has to do with female rank within the particular society, suggesting that it is probably not responsible for variation across countries. However, in order to control for this, Column 3 in Table 7 includes a control for maternal education relative to her cohort. Although the coefficient is significant and positive (more educated women have more male children), there is no effect on the coefficient on HBV prevalence.

Finally, the presence of a partner in the household has been hypothesized to play a role in offspring sex ratio. Norberg (2004) presents extensive evidence that women who live with a spouse or partner are more likely to have male children than those who do not. If women from high hepatitis regions are more likely to be conceive children while not cohabitating then this could bias the coefficients.[20] Marital status is reported in the IPUMS, and Column 4 of Table 7 includes a dummy for married or single. While the coefficient on

---

[20]Obviously cohabitation status at conception and after birth are not perfectly correlated. This would only affect the coefficient on HBV if the breakup or consummation of marriage is related to the country of origin HBV status, which seems unlikely.

marital status has the expected sign, it is not significant and the coefficient on HBV prevalence is unchanged.

As a final note, it is worth briefly considering the issue of Africa. In general, although there are a number of North African countries included in the analysis above, there are none from Sub-Saharan Africa. This is due to data limitations: data on births in Africa are virtually non-existent, and there are limited data on HBV prevalence. However, to the extent that it is possible to tell, it seems that Africa has relatively low sex ratios and relatively high HBV prevalence. This seems inconsistent with the overall cross-country relationship; however, it may simply reflect simply an Africa-specific fixed effect. A number of authors have argued that for various reasons (much lower income levels, hormonal differences, etc), Africa has lower average sex ratios at birth than the rest of the world (James, 1987a; James, 1987b; Andersson and Bergstrom, 1998).

This is somewhat unsatisfying; it would be more promising if there was evidence of this positive relationship within Africa. In addition, it seems important to try to include Africa in the overall analysis, even if just as a robustness check. Given the absence of birth registration data from Africa, it is necessary to look elsewhere for information on sex ratios at birth across countries. I have compiled household survey data (from the Demographic and Health Surveys) on 18 countries in Sub-Saharan Africa. Women in the surveys are asked about how many children of each gender they had (including those who died or no longer live at home). Using these data it is possible to create measures of sex ratios at birth across countries. Obviously these data suffer from recollection bias and the sample sizes are relatively small (usually only a few thousand women per country). The recollection bias is particularly relevant if we think that women are more likely to forget children who die very early, given that male children are more likely to die in infancy. Regardless, the comparison across countries for which these survey data are available should be consistent. The cross country comparison within Africa is presented in Column 1 of Table 8. The coefficient on hepatitis B prevalence in Africa is 0.00077. This is actually quite close in magnitude to the estimate from Table 6. In addition, the coefficient is significant at the 5% level, even with a very small sample size.

This alone provides some comfort on the issue of Africa. However, it is also worth

considering what happens when the data on Africa are included with the data from other countries. Given that there is some evidence of genetic differences producing lower sex ratios in Africa, and that the data source is very different, it is not obvious exactly how this regression should be run. In Columns 2 and 3 of Table 8 I show two specifications. In Column 2 I simply include the African countries in the cross-country regression and include the same controls as in Table 6. The coefficient on hepatitis is obviously smaller here than in the regressions shown in the paper, but it is not enormously different, and it is still significant at the 1% level. In Column 3 I also include regional fixed effects (this mimics Column 4 of Table 6). The magnitude in this case is close to the effect in Table 6 and highly significant. In general, I think these results provide some comfort that Africa is not a particular challenge to the robustness of the HBV-sex ratio connection.

4.5   Calibrating the Effects of HBV on Sex Ratio

As outlined in Section 2, the effect of HBV on sex ratio at birth is estimated from the individual level data. I have created a dataset in which each child observed in scientific studies is an observation, with a gender and an indicator for parental HBV carrier status. I also include study-specific fixed effects. There are two concerns that are worth raising here. The first is that, as discussed in section 4.2, there are data quality issues in the individual-level studies. The sex ratio among non-carriers in the Greece 1 study is very high, and it is very low in the Kar Kar Island study. Chahnazarian et al (1988) have suggested that this may be due to sample selection. The use of study specific fixed effects will partially avoid this; essentially, this technique assumes that the shift produced by hepatitis B is the same in all studies, but the level may be different. In other words, the base on which the hepatitis B effect operates is higher in the Greece 1 study than in the Kar Kar Island study. As long as the sample selection operates similarly on HBV carrier and non-carrier parents, this methodology will actually lead to a *downward* bias in the estimate of the effect of HBV on sex ratio.[21] Given that HBV status is unknown to the individuals, and there will be no

---

[21]It is easy to see why this is in a world where the sample is selected on having girls (as in the Kar Kar Island case). Imagine that all families have three children, but only families with at least one girl are tested. This biases the sex ratio downward more for carriers, since those families are more likely to have three boys. It may be less clear in a world in which the sample is selected on having at least one boy (as might be the

manifestation during this period of the individual's life, the assumption that the selection operates similarly on both groups does not seem unreasonable.

It is worth noting, however, that if the Greek sample is biased because of under-reporting of girls (rather than because of sample selection), this could lead to an upward bias in the estimates of the effect of HBV on sex ratio (the Kar Kar Island study would still produce a downward bias even if the sample is biased due to under-reporting of boys). Eliminating the Greece 1 sample from the estimation makes little difference – the estimate of offspring sex ratio to carriers moves from 1.54 boys for each girl to 1.52 boys for each girl.

A second concern is with the standard errors. In most cases, in the individual level data, there are multiple children per family in the study. Given this, it is arguably appropriate to cluster by family in the regression. Unfortunately, it is not possible to do this in the regression, since I do not observe which child is in which family. Instead, I will directly adjust the standard errors using the adjustment formula in Moulton (1990) and Kloek (1981). Details are in Appendix C. It is worth noting that the adjustment makes almost no difference. The standard error on the hepatitis B coefficient before adjustment is 0.0225 and after is 0.0226. The reason for this is simple. The degree to which the standard errors are biased down by neglecting to cluster depends on the correlation within the observations; in this case, it depends on the correlation of gender across children within a household. This correlation is almost zero. If the first child is a boy, the second child is generally not more likely to be a boy. For this reason, the adjustment for clustering is minimal.

The estimate of the effect of hepatitis B on sex ratio is presented in the upper panel of Table 9. In addition, I calculate the effects of HBV on sex ratio that are implied by the aggregate data using the methodology outlined in Equation (4), and these estimates are reported in the lower panel of Table 9. The estimate of the effect of HBV on sex ratio is 1.54, implying that HBV carriers have 1.54 boys for each girl. The 95% confidence interval is relatively large, from 1.29 to 1.87, due largely to the small samples. However, it is worth noting even the lower bound is far above the average sex ratio for non-carriers. The estimates derived from the aggregate data are consistent with the individual-level estimates. The cross

---

case in a situation like the Greece 1 study). However, in this case the sex ratio is biased upwards more for the non-carriers, since non-carrier families are more likely to have three girls, so the share of the total girls who are not reported is higher for non-carriers.

country estimates imply a sex ratio at birth for HBV carriers of between 1.40 and 1.60, very close to the ratio implied by the individual level data. The ratio implied by the Alaskan data is somewhat higher.

Using the estimates from Table 9, it is possible to calculate the predicted sex ratio at birth using data on HBV prevalence in the missing women countries. In this case, I use prevalence for pregnant women only (where available). Table 10 presents estimates of HBV prevalence, predicted sex ratio at birth and the range of empirical estimates of sex ratio at birth for each country.[22] The predicted sex ratios at birth are all well inside the range of empirical estimates for each country.[23] These results suggest that at least some of the differences in sex ratio at birth displayed in these countries can be explained by variation in HBV rate.

5   Estimating the Missing Women

In the first subsection here, I use the estimates from previous sections to calculate the share of the missing women that are explained by Hepatitis B. In the second subsection, I discuss an alternative counterfactual derived from cross-regional relationships in China and India.

5.1   Baseline Estimates

The HBV-adjusted number of missing women is calculated exactly as in Coale (1991), except that in the new estimates I use the sex ratio at birth predicted by HBV rate in Table 10, rather than a constant of 1.059 for each country. The mortality patterns assumed after birth are the same as those used in Coale (1991) and reflect mortality estimates from the model life tables in Coale, Demeny and Vaughan (1983).[24]

---

[22]The predicted sex ratios assume that carriers have the sex ratio estimated in Table 9 (1.54) and non-carriers have a sex ratio of 1.048, which is the sex ratio at birth in the countries in the cross-country sample with the lowest hepatitis B prevalence.

[23]For all countries but India and Nepal, the range represents the range of sex ratios across years in the United Nations Demographic Yearbook; for India and Nepal it represents the range by maternal age for women in household surveys. In the case of Nepal there are relatively few women of any given age in the surveys, which explains the unusually large range.

[24]These life tables are widely used by demographers. They use historical data on population in a number of regions to calculate population projections given any particular life expectancy and gross reproductive rate. The tables can be easily used to calculate expected sex ratios by gender, which are adjusted for sex ratio at

The new estimation appears in Table 11. In Panel A, I compare the HBV-adjusted estimates to those of Coale (1991). After adjusting for HBV, roughly 47%, or 31 million, of the Coale (1991) missing women are explained. The country-by-country breakdown reveals large variation in the percent of the puzzle that is accounted for by HBV. In particular, the high HBV rate in China explains about 75% of the missing women in that country. In contrast, in India, Bangladesh, Pakistan and Nepal less than 20% of the puzzle is explained.[25] The numbers for Bangladesh, Egypt and Turkey/Syria are intermediate. The 95% confidence interval on share explained is fairly large – roughly 13.5% to 70% – but it is worth noting that in China even the lower bound suggests around 30% of the missing women are accounted for.

These results are broadly consistent with the evidence presented in Section 3. India and Pakistan present sex ratio patterns by age that clearly display excess female mortality among young children. In China, however, the pattern of mortality and the evidence on historical sex ratios supports the possibility that HBV-produced differences in sex ratio at birth are largely the root of the gender disparity.

In Panel B of Table 11 I compare the new estimates to the original estimates from Sen (1992). In his original work, Sen assumed that sex ratios from Sub-Saharan Africa would prevail in Asia absent intervention – in particular, he assumed the sex ratio would be 0.977. As can be seen in Table 11, relative to that estimate the HBV adjustment accounts for 70% of the puzzle, with a 95% confidence interval ranging from 50% to 85%. More accurately, combining both realistic mortality rates (as in Coale, 1991) *and* adjusting for the HBV-determined sex ratio at birth results in an estimate of 32 million missing women.[26]

---

birth to give expected sex ratio overall.

[25]The results suggest that the number of missing women in Nepal actually increases after adjusting for HBV. This is due to the fact that in the original Coale(1991) estimates the assumed sex ratio at birth is actually quite high (1.059). Although the Coale(1991) ratio is lower than the HBV-adjusted sex ratio for most countries in this region, Nepal has a very low HBV rate, so the predicted sex ratio at birth is only 1.054.

[26]More recently, Klasen and Wink (2002) have updated the number of missing women to reflect sex ratios and population sizes in the late 1990s. They use a slightly different set of life tables than in Coale(1991) and do not use a fixed sex ratio at birth. They estimate a total of 93 million missing women. The comparison with their estimates yields similar results to the others: HBV explains about 55% of the missing women.

5.2 An Alternative Counterfactual

As mentioned in the introduction, the estimates above assume that, in the absence of hepatitis B, there would be no compensating behavior change affecting sex ratios. This is likely to be unrealistic. Given that even the least conservative estimates in this paper suggests a significant role for discrimination in the sex ratio imbalance, it seems extremely likely that some of the extra girls born in the absence of hepatitis would die prematurely due to discrimination. In this sense, the estimates above present the total effect of hepatitis on sex ratios, and we may be more interested in the partial effect. Calculating this requires a counterfactual world in which gender preference is the same, but hepatitis rates differ (as opposed to the current counterfactual – the West – which has lower levels of gender bias *and* lower levels of hepatitis B).[27] One possible way to approach this is to explore cross-country variation in hepatitis rates within the missing women countries. This is additionally important, as a cross-regional relationship between hepatitis and sex ratio at birth would strengthen the case for a relationship.

The theory here is simple, although requires some significant assumptions: if we are willing to assume that average gender preference is constant across regions within a country, then the relationship between hepatitis and sex ratios across regions should help capture the partial effect. Said differently, the sex ratio in the regions with little or no hepatitis should reflect the sex ratio that would prevail everywhere in the absence of hepatitis. To explore this, I make use of some (limited) data on cross-regional variation in India and China. I will focus in particular on the Chinese case, since the data in India are even more limited, and the baseline estimates suggest that HBV is substantially less important in explaining the gender imbalance there.

For India, estimates of HBV rate are available from eight states (World Health Organization, 2002) and these are matched with data on the share of boys among children ages 0-6 from Dreze and Sen(2003). In China, one study has done HBV testing in 60 of the roughly 2300 Chinese counties (Campbell et al, 1990), although only about 100 people were

---

[27]It is worth noting that while intuitively the counterfactual is the West, the true counterfactual is the sex ratio at birth assumed in Coale (1991). This is 1.059, well above the estimate of around 1.050 that is generally assumed in Western countries. In this sense there has already been some upward adjustment of the sex ratio to reflect behavior in these countries.

tested in each county. These HBV rates are matched with births from 1989 and 1990, as reported in the 1989 Chinese population census. There are, of course, reliability issues associated with all of these data. In addition, even in China, they cover only a small subset of regions, implying that the analysis here should be viewed as suggestive.

Table 12 shows a simple regression of sex ratio on HBV across region in India (Column 1) and China (Columns 2). In both cases the relationship is positive. In the case of India the relationship is significant at the 10% level (although with only eight observations it is difficult to draw conclusions). In China, however, this relationship is significant at the 1% level (the significance of relationship is not driven by outliers – it remains strongly significant in quantile and outlier-robust regressions). In addition, in Column 3 I show the regression for China including controls for the share of the population in agriculture and the income in the province. The point estimate is largely unaffected, and it is still statistically significant. It is worth noting that the size of the hepatitis effect in the Chinese data is similar to that in the overall cross-country regressions.

Although the estimates in Table 12 support the hepatitis-sex ratio connection, and suggest that hepatitis explains some of the cross-regional variation in sex ratios, they do not directly address the counterfactual issue. To do this, it is necessary to get an estimate of the sex ratio in China without HBV. There are two ways to do this – either extrapolate from the regression coefficients in Column 3 of Table 12, or simply observe the sex ratios in low-HBV regions in China. Extrapolating from the regression gives a non-HBV sex ratio at birth of 1.066; using the average sex ratio in counties with low (less than 5%) levels of hepatitis gives 1.059. These are a bit higher than the sex ratios at birth in the West. However, 1.059 is the sex ratio at birth used by Coale(1991) to calculate the number of missing women. As a result, if that is the best estimate of the non-HBV sex ratio in China, then the partial estimate of hepatitis is the same as the total (at least in China), and is given in Table 11.

If, instead, the correct counterfactual is 1.066, then the partial effect of hepatitis is less than the total effect. To calculate the partial effect we compare the number of missing women implied by the HBV-adjusted sex ratio at birth to the number of missing women implied by a sex ratio at birth of 1.066. Doing this for China indicates that the partial effect of hepatitis is to explain 18 million missing women, or around 62% of those estimated by

Coale (1991) (in contrast to the 75% that represents the total effect).

Clearly this methodology for creating the counterfactual world is not perfect, and the data are quite limited. Nevertheless, the results here suggest that while the partial effect of hepatitis may be slightly lower than the total effect due to compensating behaviors, it is still reasonable to think it is larger than zero.

6   Discussion and Conclusion

The results above clearly have implications for the missing women puzzle. Introducing heterogeneity in hepatitis B rates into the model decreases the number of missing women by around one-half. In addition, after adjusting for different rates of hepatitis B, the problem of the missing women may be more heavily focused on India and Pakistan. This focus may help target policy interventions and better understand the dynamics of this important issue.

This paper largely focuses on the mechanical connection between HBV and sex ratios. It is interesting to consider, however, that there may be scope for interactions between parental preferences and HBV (other than the compensating behavior issue). In particular, the combination of a gender-biased "stopping rule" for fertility and high prevalence of HBV could have important implications for population growth and gender balance within families.[28] A gender-biased stopping rule in a homogenous population will not alter the overall population sex ratio. However, the same stopping rule in a population affected by HBV can both mitigate the effect of HBV on sex ratios and affect population growth.

Imagine that 20% of the population are carriers of HBV and parents use a male-biased stopping rule in which they stop after the first boy. Then, in the first birth, 60% of the HBV carriers and 51% of the non-carriers will give birth to boys. This implies that more people stop having children after the first birth than would be implied by a non-HBV model. However, the population of people going on to have a second birth will contain only 16% HBV carriers. This means that, on average, later births will be more female biased. The overall interaction has two effects: slower population growth rate and more all-female families

---

[28]A gender-biased stopping rule is a general description of a fertility pattern in which parents continue to have children until they have the desired number of boys or girls and then stop.

than would be predicted by the overall sex ratio at birth. In addition, if this is an important effect, the overall sex ratio in the population will be lower than would be produced naturally by HBV, since fertility responds endogenously.

Interestingly, this interaction seems to be more than just a theoretical possibility. In particular, in some of the individual-level data on HBV and sex ratios, it is possible to see the effects of what appears to be exactly this interplay between a stopping rule and HBV infection. For the case of Plati, Greece – where the effect of HBV on sex ratio was first noted – London et al (1982) report the number of boys and girls born to carrier and non-carrier parents. The data indicate a roughly similar number of boys born to each type (1.8 boys for carrier families, 1.6 for non-carrier) but differences in the number of girls (0.7 for carrier families, 1.3 for non-carriers).

Given the sex ratios reported for carrier and non-carrier families, the observed data are quite close to what one would expect with a simple stopping rule: stop when you have two boys or three children. This rule predicts around 1.7 boys and 0.9 girls for carrier families; 1.4 boys and 1.3 girls for non-carrier families. Given that there is obviously additional individual-level heterogeneity and imperfect fertility control, this seems like a very good fit to the data. It is also worth noting that we see both the population growth rate effect and the overall sex ratio effect in this population: the average number of children per family is 2.9, rather than the 3 that prevails for the non-HBV carriers alone and the overall population sex ratio is 1.16, rather than the 1.20 that would be predicted by the overall HBV rate.

In general, this discussion and evidence suggests that there may well be scope for exploring the relationship between preferences and biological heterogeneity in fertility outcomes. To the extent that economists have used gender-biased stopping rules to explore the relationship between family size and outcomes (see, for example, Jensen, 2005) the heterogeneity introduced by HBV may provide additional insights.

Among other things, the results in this paper suggest that HBV vaccination may have effects on future sex ratios. Already the vaccine is widely administered across the world and, as the estimates in Section 4.3 suggest, the effects can be seen already. As fewer and fewer carriers are giving birth, naturally occurring sex ratios in these countries may continue to decline. With a strong tradition of son preference, one might expect even more sex-selective

abortion and other behavioral responses that increase the probability of a male child.

The introduction of the vaccine has implications outside of these countries, as well. A number of European countries are also likely to also see some decrease in their sex ratio fewer women of childbearing age are HBV carriers. If one takes seriously the theory in Becker(1981) and the evidence presented in Angrist(2002) and Chiappori et al. (2002) then these changes may have consequences for marriage markets and labor market outcomes in these countries.

The missing women problem has historically been understood as arising from a difference in preferences – some countries value women less than others, and in these places women are mistreated to the point of excess mortality. While there can be no doubt that gender bias plays a significant role in this gender imbalance, the results here suggest that differences in biology may play a role as well. Understanding both mechanisms, and their interaction, seems crucial to moving forward on this important problem.

References

Acemoglu, Daron, David Autor, and David Lyle, "Women, War and Wages: The effect of Female Labor Supply on the Wage Structure at Mid-Century," *Journal of Political Economy*, 2004, *112* (3), 497–551.

Adair, R and O Nwaneri, "Communicable disease in African immigrants in Minneapolis," *Archives of Internal Medicine*, 1999, *159* (1), 83–85.

Agnihotri, Satish, "Inferring Gender Bias from Mortality Data: A Discussion Note," *Journal of Development Studies*, 1999, *35* (4), 175–200.

Amini, S, MF Mahmoodi, S Andalibi, and AA Solati, "Seroepidemiology of hepatitis B, delta and human immunodeficiency virus infections in Hamadan province, Iran: a population based study," *Journal of Tropical Medicine and Hygine*, 1993, *96* (5), 277–287.

Andersson, R and S Bergstrom, "Is maternal malnutrition associated with a low sex ratio at birth?," *Human Biology*, 1998, *70* (6), 1101–1106.

Andrade, Dahir Ramos Junior and Dahir Ramos Andrade, "The Influence of the Human Genome on Chronic Viral Hepatitis Outcome," *Review of the Institute of Tropical Medicine*, 2004.

Angrist, Josh, "How do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation," *Quarterly Journal of Economics*, 2002, *117* (3), 997–1038.

Becker, Gary, *A Treatise on the Family*, Cambridge, MA: Harvard University Press, 1981.

Blumberg, Baruch, *Hepatitis B: The Hunt for a Killer Virus*, Princeton: Princeton University Press, 2002.

Campbell, T. Colin, Junshi Chen, Chongbo Liu, Junyao Li, and Banoo Parpia, "Non-association of Aflatoxin with Primary Liver Cancer in a Cross-Sectional Ecological Survey in the People's Republic of China," *Cancer Research*, 1990, *50*, 6882–6893.

Cazal, P, JM Lemiare, and M Robinet-Levy, "Hepatitis B et Rapport du Masculinite," *Revue franaise de transfusion et immunohmatologie*, 1976, *19*, 577–581.

Chahnazarian, A, BS Blumberg, and WT London, "Hepatitis B and the sex ratio at birth: a comparative analysis of four populations," *Journal of Biosocial Science*, 1988, *20* (3), 357–370.

Chambliss, Rollin, "The Geographic Factor in the Human Sex Ratio at Birth," *Social Forces*, 1949, *28* (2), 190–195.

Chen, Ta, "Sex, Age, Size of Family and Density of Population," *American Journal of Sociology*, 1947, *52*, 17–24.

Chiao, Chi-Ming, "A Study of Chinese Population," *Milbank Memorial Fund Quarterly Bulletin*, 1934, *12* (2), 171–184.

Chiappori, Pierre-Andre, Bernard Fortin, and Guy Lacroix, "Marriage Market, Divorce Legislation and Household Labor Supply," *Journal of Political Economy*, 2002, *110* (11), 37–72.

Coale, Ansley, "Excess Female Mortality and the Balance of the Sexes: An Estimate of the Number of 'Missing Females'," *Population and Development Review*, 1991, *17*, 517–523.

⎯ **and Judith Banister**, "Five Decades of Missing Females in China," *Demography*, 1994, *31* (3), 459–479.

⎯ **, Paul Demeny, and Barbara Vaughan**, *Regional Model Life Tables and Stable Populations*, Princeton: Princeton University Press, 1983.

Department of Vaccines, "Introduction of Hepatitis B Vaccine into Childhood Immunization Services," Technical Report, World Health Organization 2001.

Drew, JS, BS Blumberg, and J Robert-Lamblin, "Hepatitis B virus and sex ratio of offspring in East Greenland," *Human Biology*, 1986, *58* (1), 115–120.

⎯ **, WT London, BS Blumberg, and S Serjeanston**, "Hepatitis B virus and sex ratio on Kar Kar Island," *Human Biology*, 1982, *54* (1), 123–135.

⎯ **, ⎯ , ET Lustbader, JE Hesser, and BS Blumberg**, "Hepatitis B Virus and the Sex Ratio of Offspring," *Science*, 1978, *201* (4357), 687–692.

Dreze, Jean and Amartya Sen, *India: Development and Participation*, 2nd ed., Oxford: Oxford University Press, 2003.

Duong, Tran Nhu, Norio Horiike, Kojiro Michitaka, Chen Yan et al., "Comparison of Genotypes C and D of the Hepatitis B Virus in Japan: A Clinical and Molecular Biological Study," *Journal of Medical Virology*, 2004, *72*, 551–557.

Edlund, Lena, "Son Preference, Sex Ratios and Marriage Patterns," *Journal of Political Economy*, 1999, *107* (6), 1275–1304.

Gjerdingen, DK and V Lor, "Hepatitis B status of Hmong patients," *Journal of the American Board of Family Practitioners*, 1997, *10* (5), 322–328.

Goldin, Claudia, "The Role of World War II in the Rise in Women's Employment," *American Economic Review*, 1991, *81* (4), 741–756.

Harpaz, Rafael, Brian McMahon, Harold Margolis et al., "Elimination of New Chronic Hepatitis B Virus Infections: Results of the Alaska Immunization Program," *Journal of Infectious Disease*, 2000, *181*, 413–418.

Hayes, EB, SB Talbot, ES Matheson, HM Pressler, AB Hanna, and CA McCarthy, "Health status of pediatric refugees in Portland, ME," *Archives of Pediatric and Adolescent Medicine*, 1998, *152* (6), 564–568.

Hesser, JE, J Economidou, and BS Blumberg, "Hepatitis B surface antigen (Australia antigen) in parents and sex ratio of offspring in a Greek population," *Human Biology*, 1975, *47* (4), 415–425.

Hsu, HM, CF Lu, SC Lee, SR Lin, and DS Chen DS, "Seroepidemiologic survey for hepatitis B virus infection in Taiwan: the effect of hepatitis B mass immunization," *Journal of Infectious Disease*, 1999, *179* (2), 367–370.

Hull, Terence, "Recent Trends in Sex Ratio at Birth in China," *Population and Development Review*, 1990, *16* (1), 63–83.

James, William, "The Human Sex Ratio. Part 1: A review of the literature," *Human Biology*, 1987, *59* (5), 721–752.

___ , "The Human Sex Ratio. Part 2: A hypothesis and a program of research," *Human Biology*, 1987, *59* (6), 873–900.

Jensen, Robert, "Equal Treatment, Unequal Outcomes? Generating Sex Inequality Through Fertility Behavior.," 2005. Harvard University.

Junhong, Chu, "Prenatal Sex Determination and Sex Selective Abortion in Rural Central China," *Population and Development Review*, 2001, *27* (2), 259–281.

Kent, Heather, "BCs Chinese migrants a healthy lot, MDs find," *Journal of the Canadian Medical Association*, 2000, *162* (2), 256.

Klasen, Stephan, "Missing Women Reconsidered," *World Development*, 1994, *22* (7), 1061–1071.

___ , "Nutrition, Health and Mortality in Sub-Saharan Africa: Is there a Gender Bias?," *Journal of Development Studies*, 1996, *32* (6), 913–932.

___ **and Claudia Wink**, "A Turning Point in Gender Bias in Mortality? An Update on the Number of Missing Women," *Population and Development Review*, 2002, *28* (2), 285–312.

Kloek, T., "OLS Estimation in a Model Where a Microvariable is Explaiend by Aggregates and Contemporaneous Disturbances are Equicorrelated," *Econometrica*, 1981, *49* (1), 205–207.

Kobayashi, Mariko, Yasuji Arase, Kenji Ikeda, Akihito Tsubota et al., "Clinical features of hepatitis B virus genotype A in Japanese patients," *Journal of Gastroenterology*, 2003, *38*, 656–662.

Kojima, Fuhito, "Economics of Infanticide," 2005. Mimeo, Harvard University.

Lazarus, John, *Sex Ratios: Concepts and Research Methods*, Cambridge University Press, 2002.

Lee, DH, JH Kim, JJ Nam, HR Kim, and HR Shin, "Epidemiological findings of hepatitis B infection based on 1998 National Health and Nutrition Survey in Korea," *Journal of Korean Medical Science*, 2002, *17* (4), 457–462.

Lin, HH, LY Wang, CT Hu, SC Huang, LC Huang, SS Lin, YM Chiang, TT Liu, and CL Chen, "Decline of hepatitis B carrier rate in vaccinated and unvaccinated subjects: sixteen years after newborn vaccination program in Taiwan," *Journal of Medical Virology*, 2003, *69* (4), 471–474.

Livadas, D, J Economidou, JE Hesser, DA Koutras, and S Hadziyannis, "Fertility and Sex Ratio of Offspring of Female HBsAg Carriers," *Journal of the Royal Society of Medicine*, 1979, *72*, 509–512.

London, W. Thomas, Richard Stevens, Frances Shofer, Jean Drew, Joann Brunhofer, and Baruch Blumberg, *Effects of Hepatitis B Virus on the Mortality, Fertility and Sex Ratio of Human Populations*, Philadelphia, PA: The Franklin Institute Press, 1982.

Mazzur, S and TM Watson, "Excess males among siblings of Australian antigen carriers," *Nature*, 1974, *250* (461), 60–61.

McMahon, BJ, S Schoenberg, L Bulkow, RB Wainwright, MA Fitzgerald, AJ Parkinson, E Coker, and D Ritter, "Seroprevalence of hepatitis B viral markers in 52,000 Alaska Natives.," *American Journal of Epidemiology*, 1993, *138* (7), 544–549.

Miyakawa, Yuzo and Masashi Mizokami, "Classifying Hepatitis B Virus Genotypes," *Intervirology*, 2003, *46*, 329–338.

Moulton, Brent, "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables in Micro Units," *The Review of Economics and Statistics*, 1990, *72* (2), 334–338.

Ni, YH, MH Chang, LM Huang, HL Chen, HY Hsu, TY Chiu, KS Tsai, and DS Chen, "Hepatitis B virus infection in children and adolescents in a hyperendemic area: 15 years after mass hepatitis B vaccination," *Annals of Internal Medicine*, 2001, *135* (9), 796–800.

Norberg, Karen, "Partnership Status and the Human Sex Ratio at Birth," *Proceedings of the Royal Society B*, 2004, *Forthcoming*.

Patel, PA and MD Voigt, "Prevalence and interaction of hepatitis B and latent tuberculosis in Vietnamese immigrants to the United States," *American Journal of Gastroenterology*, 2002, *97* (5), 1198–1203.

Qian, Nancy, "Missing Women and the Price of Tea in China: The Effect of Relative Female Income on Sex Imbalance," 2005. MIT Working Paper.

Rahman, M, Amanullah, H Sattar, M Rahman, HA Rashid, and AS Mollah, "Seroepidemiological study of hepatitis B virus infection in a village," *Bangladesh Medical Research Council Bulletin*, 1997, *23* (2), 38–41.

Ranger-Rogez, Sylvie and Francois Denis, "Hepatitis B Mother to Child Transmission," *Expert Review of Anti-Infective Therapies*, 2004, *2* (1), 133–145.

Republic of China, *Taiwan-Fukien Demographic Factbook*, Minguo: Republic of China, 1974-2002,annual.

Rosenzweig, Mark and T. Paul Shultz, "Market Opportunities, Genetic Endowments and Intrafamily Resource Distribution: Child Survival in India," *American Economic Review*, 1982, *72* (4), 803–815.

Sen, Amartya, "Missing Women," *British Medical Journal*, 1992, *304*, 587–588.

_____ , "Mortality as an Indicator of Economic Success and Failure," *Economic Journal*, 1998, *108*, 1–25.

_____ , "More than 100 Million Women are Missing," *New York Review of Books*, December 20, 1990.

She, SL, LY Shi, YJ Wu, ZZ Li, CZ Zheng, YP Wu, and XH Yu XH, "A seroepidemiologic study of hepatitis B virus infection among barbers in Huangshi City, Hubei, China," *Microbiology and Immunology*, 1988, *32* (2), 229–233.

Simmonds, P, "Reconstructing the Origins of Human Hepatitis Viruses," *Philosophical Transactions of the Royal Society B*, 2001, *356*, 1013–1026.

Trivers, Robert and Dan Willard, "Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring," *Science*, 1973, *179* (4068), 90–92.

United Nations, "Demographic Yearbook Historical Supplement," Technical Report, United Nations 1997.

_____ , "Demographic Yearbook," Technical Report, United Nations 2001.

Wang, CS, TT Chang, WJ Yao, and P Chou, "Comparison of hepatitis B virus and hepatitis C virus prevalence and risk factors in a community-based study," *American Journal of Tropical Medicine and Hygiene*, 2002, *66* (4), 389–393.

Wang, Z, J Zhang, H Yang, X Li, S Wen, Y Guo, J Sun, and J Hou, "Quantitative analysis of HBV DNA level and HBeAg titer in hepatitis B surface antigen positive mothers and their babies: HBeAg passage through the placenta and the rate of decay in babies," *Journal of Medical Virology*, 2003, *71* (3), 360–366.

World Health Organization, "Prevention of Hepatitis B in India: An Overview," Technical Report, World Health Organization, Regional Office for South-East Asia 2002.

Yi, Zeng, Tu Ping, Gu Baochang, Xu Yi, Li Bohua, and Li Yongping, "Causes and Implications of the Recent Increase in Reported Sex Ratio at Birth in China," *Population and Development Review*, 1993, *19* (2), 283–302.

Zhuo, J, G Tao, SH Ebrahim, S Wang, Z Luo, and H Wang, "The relationship of hepatitis B virus infection between adults and their children in Guangxi Province, China," *Journal of Hepatology*, 2000, *33* (4), 628–31.

Appendix A: Cross-Space and Cross-Time Variation

This appendix addresses issues of cross-space and cross-time variation in HBV carrier prevalence. I first present a simple model of viral transmission over time. I then calibrate the model for three areas and demonstrate the importance of infection rates at different ages in the overall carrier prevalence. Finally, I discuss the steady state properties of virus prevalence.

Assume there are three age groups – infant, child and adult. In each group there is a share infected (a share who are carriers) and a share uninfected (those who are not carriers). I therefore have six groups, and can write the existing infection profile as a 6×1 vector. Denote μ as the carrier rate among infants, γ as the carrier rate among children and δ as the carrier rate among adults. The vector describing the current state of the system can therefore be written as:

$$\begin{bmatrix} \text{infant carrier rate} \\ \text{uninfected infant rate} \\ \text{infected children rate} \\ \text{uninfected children rate} \\ \text{infected adult rate} \\ \text{uninfected adult rate} \end{bmatrix} = \begin{bmatrix} \mu \\ 1-\mu \\ \gamma \\ 1-\gamma \\ \delta \\ 1-\delta \end{bmatrix}$$

It is necessary to define a transition process for the system. Very generally, this will be a 6×6 matrix that maps infection rates this period into infection rates in the next period. Note that the matrix is restricted by the fact that infant infection rates map into child infection rates, child infection rates map into adult infection rates and adult infection rates map into infant infection rates (through mother to child transmission). The most general form of the transition matrix is:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & a & b \\ 0 & 0 & 0 & 0 & 1-a & 1-b \\ c & d & 0 & 0 & 0 & 0 \\ 1-c & 1-d & 0 & 0 & 0 & 0 \\ 0 & 0 & e & f & 0 & 0 \\ 0 & 0 & 1-e & 1-f & 0 & 0 \end{bmatrix}$$

The matrix can be simplified still further by the assumption (consistent with the disease) that once infected an individual stays infected for their entire life. This means that $c = 1$ (all infected infants become infected children) and $e = 1$ (all infected children become infected adults). In addition, I note that uninfected women cannot transmit the virus to their children. This implies that $b = 0$. In order to calibrate the model, however, it will be helpful to simplify still further and separate the remaining parameters a, d and f into their components. In particular, it will be useful to differentiate between the chance of becoming infected and the chance of becoming a carrier given that infection takes place.

The chance of becoming a carrier of the virus given infection varies by age. The typical parameters cited in the literature are 90-95% chance of becoming a carrier conditional on infection for infants, 25-50% for very young children, 20% for older children and 2-10% for adults (Department of Vaccines, 2001; Ranger-Rogez and Denis, 2004). In this exercise, I will assume 90% for infants, 30% for all children, and 5% for adults. Denote the chance of

infection as an infant as x, the chance of infection as a young child as y and the chance of infection as an adult as z.[29] The matrix A above can now be rewritten as:

$$
A = \begin{bmatrix}
0 & 0 & 0 & 0 & 0.9x & 0 \\
0 & 0 & 0 & 0 & 1 - 0.9x & 1 \\
1 & 0.3y & 0 & 0 & 0 & 0 \\
0 & 1 - 0.3y & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0.05z & 0 & 0 \\
0 & 0 & 0 & 1 - 0.05z & 0 & 0
\end{bmatrix}
$$

The first two rows show the transition from adult infection rate to infant infection rate. The second row shows transition from infant to child – all infected infants are infected children, and there is an additional possibility of infection and carrier state as child. A similar logic follows for adults. To learn the steady state for the epidemic, I simply consider the limit of $(A^n) v$ as $n \rightarrow \infty$, where v is the initial state of the epidemic. I note that in this type of transition matrix stability implies that the initial starting values do not impact the final steady state.

The structure of the transition matrix should make it clear that infections among infants and children are far more important in the steady state of the epidemic than infections among adults. Since only 5% of adults who are infected become carriers, the vast majority of carriers in a given area will be infected as infants or as children. This is consistent with the evidence that sexual behavior within a country (the primary method of infection for adults) is not an important determinant of HBV carrier rate.

Examples of Transmission

I now provide some estimates of x, y and z for two areas (China and Alaska) in an effort to provides some insight into why HBV prevalence varies across theses areas. x is the simplest parameter to estimate. The chance of maternal transmission to an infant is largely a function of whether the mother is currently replicating the virus. For mothers who are replicating the virus, the chance of infecting the infant is around 90%; for those who are not replicating, the chance is 10-30% (Ranger-Rogez and Denis, 2004). Here, I will assume that the chance of passing the virus for non-replicating women is 20%, and use estimates about the share of those who are replicating in each area to estimate x (it is worth noting that whether a woman is actively replicating the virus seems to be largely determined by which genotype of HBV she has – those with genotype C, for example, are likely to be replicating the virus, whereas those with genotype D or A are much less likely).

Estimating y and z is more complex. Fortunately for this exercise the virus leaves its mark even in those who clear it from their system. In particular, those who have been infected but are not carriers will nevertheless test positive for the antibody to HBV (anti-HBs). It is therefore possible to use information about the prevalence of anti-HBs by age to determine the infection rate.

---

[29]It is worth noting that the chance of a child becoming infected is a function of the share of infants that are infected (since often infection will be from other children of the individual's age group). For simplicity, I abstract from this here, while recognizing that this fact means that the importance of the infection rate among infants is understated.

Guangxi Province, China

Calibrations:

x : Around half of HBV carriers in China actively replicate the virus (Wang et al, 2003; Zhuo et al, 2000). This suggests that the transmission rate to infants is $x = 0.55$.

y : Results (from Zhuo et al, 2000) indicate that roughly 36% of children of non-infected parents have been exposed to the virus, implying that $y = 0.36$.

z : Results (from Zhuo et al, 2000) suggest that overall around 65% of adults have been exposed to the virus (but are not carriers). This implies that $z = 0.3$ (since the total exposures among adults are the sum of those exposed as children and those exposed as adults).

Given these parameter values, the stable steady state of the system is:

$$\begin{bmatrix} \text{infant carrier rate} \\ \text{infected children rate} \\ \text{infected adult rate} \end{bmatrix} = \begin{bmatrix} 0.106 \\ 0.202 \\ 0.214 \end{bmatrix}$$

The actual carrier rates among children and adults in this area are 20% and 20%, remarkably close to the stable steady state suggested above.

Alaska

All data for this calibration come from McMahon et al (1993), in which data on HBsAg and anti-HBs are reported for all Alaskan Natives.

x: The share of women who are positive for viral replication is around 23%, implying that $x = 0.36$

y : The available data on anti-HBs among young children suggests that $y = 0.06$.

z : Data suggests overall prevalence of anti-HBs is 17%, which is compared to 6% and implies that $z = 0.11$.

Given these parameter values the stable steady state is:

$$\begin{bmatrix} \text{infant carrier rate} \\ \text{infected children rate} \\ \text{infected adult rate} \end{bmatrix} = \begin{bmatrix} 0.011 \\ 0.029 \\ 0.034 \end{bmatrix}$$

This is quite close to the overall prevalence of around 3%.

Obviously the calculations here are quite loose, but they do give some sense of what influences carrier rates across space. In particular, the much higher infection rate among infants and children in that area of China seems to be largely responsible for the high infection rate there.

Cross-Time Variation and Stability

A final question is whether the speed of convergence of the system described here, combined with information about the age of the virus, is consistent with the possibility that the current infection rates are a steady state. Estimates suggest the virus is at least 3000 years old

(Simmonds, 2001). With a generation length of 30 years, this implies 100 generations. If I consider the example of China, 100 generations is sufficient to get to the steady state. In fact, in this case, stability is achieved within roughly 15 generations, and is very close within 10. At least in the case of this approximate process, convergence is relatively fast compared to the age of the virus, and the transition process is consistent with current stability.

Appendix B: Hepatitis Rate Categories

High (>8%)		Medium (2%-7%)		Low (< 2%)	
Armenia	Philippines	Albania	Romania	American Samoa	Ireland
Azerbaijan	Reunion	Algeria	Russia	Antigua	Jamaica
Bahrain	Seychelles	Bangladesh	Slovenia	Argentina	Jersey
Bolivia	Singapore	Bosnia	Spain	Australia	Latvia
Botswana	South Korea	Brazil	Sri Lanka	Austria	Lithuania
Brunei	Tadjikistan	Croatia	Suriname	Bahamas	Luxembourg
Bulgaria	Thailand	Cyprus	Syria	Belarus	Malta
Cape Verde	Togo	Dominican Rep.	Tunisia	Belgium	Martinique
China	Uzbekistan	Egypt	Turkey	Belize	Mexico
Columbia	Zimbabwe	El Salvador	UAE	Bermuda	Netherlands
Georgia		French Guiana	Venezuela	Canada	New Caledonia
Greenland		Gibraltar	Yugoslavia	Cayman Islands	New Zealand
Hong Kong		Greece	India	Channel Islands	Nicaragua
Israel		Guatemala	Nepal	Chile	North Ireland
Jordan		Guyana		Cook Islands	Norway
Kyrgyzstan		Honduras		Costa Rica	Panama
Kzakhstan		Iran		Cuba	Paraguay
Liberia		Iraq		Czech Republic	Puerto Rico
Macaco		Italy		Czechoslovakia	Samoa
Malawi		Japan		Denmark	San Marino
Malaysia		Kuwait		Dominica	Scotland
Maldives		Lybia		Ecuador	Slovakia
Mali		Macedonia		Estonia	St Lucia
Mauritius		Morocco		Fiji	Sweden
Moldova		Pakistan		Finland	Switzerland
Mongolia		Poland		France	Tonga
North Korea		Portugal		Germany	Trinidad
Peru		Qatar		Guadeloupe	Ukraine
				Guam	United Kingdom
				Hungary	United States
				Iceland	Uruguay

Appendix C: Standard Error Adjustment

Following Kloek(1981), Moulton(1990) equation (3) provides a simple formula for the adjusted OLS covariance matrix in the presence of clustered observations:

$$C = \sigma^2 (X'X)^{-1}[1 + (m-1)\rho]$$

As usual, $\sigma^2(X'X)^{-1}$ is the standard OLS covariance matrix; $[1 + (m-1)\rho]$ is the adjustment factor, with m representing the number of observations per group and ρ the inter-group correlation. This formula will hold strictly only when the number of observations per group is the same. However, as Moulton (1990) points out, it is a close approximation to the true adjustment factor as long as the groups are of similar size.

In the example used in this paper I do not observe the family size for each child directly. Instead, I observe the average family size for each study. To be conservative in the adjustment factor, I assume that m is equal to the largest average family size in the studies used – $m = 6$.

The other adjustment factor necessary here is ρ – the interclass correlation. Essentially, this is simply the correlation of gender across any two randomly chosen children in the same family. Ideally, this would be done in places with low levels of hepatitis B. To get some sense of the size of ρ, I put together household surveys from several countries with low HBV rates (Nicaragua, Dominican Republic and Guatemala). In this set of countries, I look at the correlation in gender of the first and second child, the first and third child, and the second and third child in roughly 12,000 families. This correlation is 0.0025. With this, I have sufficient information with which to adjust the standard errors, and it is the corrected standard errors that are presented in Table 9.

Table 1
Historical Estimates of Sex Ratio at Birth
Data from Coale and Banister (1994)

Year of Birth	Sex Ratio at Birth (Boys/Girls)
1936-1940	1.138
1941-1945	1.114
1946-1950	1.083
1951-1955	1.094
1956-1960	1.096
1961-1965	1.065
1966-1970	1.079
1971-1975	1.073
1976-1980	1.088
1981-1985	1.086
1985-1989	1.128
Average	**1.095**

Notes: The data in the table are taken from Coale and Banister (1994), Table 3. Their data come from a retrospective fertility survey run in China in 1982.

Table 2
Offspring Sex Ratios Immigrants and Natives in the IPUMS

Panel A: Offspring Sex Ratios among Chinese Immigrants		
Year	Number of Children	Sex Ratio (Boys/Girls)
1880-1930	45	1.250
1940-1970	1144	1.047
1980	1109	1.136
1990	1312	1.092
2000	2823	1.119
Average		**1.105**
Average, 1980-2000		**1.115**

Panel B: Offspring Sex Ratios among Immigrants from Low-HBV Countries and Native Whites

Other Immigrants, 1900-2000	**1.041**
Natives, 1900-2000	**1.044**

Notes: This table reports sex ratios among children ages 1-5 observed in the U.S. Census IPUMS 1% and 5% samples. Panel A is limited to children who were born in the United States to mothers who are Chinese immigrants. The first line of Panel B is limited to children whose mothers are immigrants from countries with a hepatitis B prevalence of less than 2%. The second line of Panel B is the average for children of Native whites. There is no data for 1890 or 1930.

<div align="center">

Table 3

Hepatitis B and Sex Ratio: Individual Level Estimates

</div>

Location	HBV Status	Sons	Daughters	Sex Ratio
Greenland	Positive	64	60	1.07
Greenland	Negative	174	194	0.90
Kar Kar Island	Positive	63	54	1.17
Kar Kar Island	Negative	163	206	0.79
Greece 1	Positive	90	51	1.77
Greece 1	Negative	287	255	1.13
Philippines	Positive	66	41	1.61
Philippines	Negative	304	301	1.01
Greece 2	Positive	52	30	1.73
Greece 2	Negative	1006	955	1.05
France	Positive	20	12	1.66
France	Negative	149	122	1.22

Notes: This table shows sex ratios among the children of carrier and non-carrier parents in four regions. Data were collected by testing married women and, in all cases except for Greenland, their husbands for HBV. Detailed reproductive histories were also collected. The table represents all births to women in these samples, with generally more than one birth to each woman. The last two studies (Greece 2 and France) were designed specifically to test the hypothesis that HBV affects offspring sex ratio, and were run after the original theory was published. The citations for each study are as follows: Greenland – Drew, Blumberg and Robert-Lamblin (1986); Kar Kar Island – Drew et al (1982); Greece 1 – Hesser, Economidou and Blumberg (1975); Philippines – Chahnazarian et al (1988); Greece 2 – Lividas et al (1979); France – Cazal, Lemaire and Robinet-Levy (1976).

Table 4[a]
Sex Ratio and HBV Vaccination: Alaskan Natives and Non-Natives
Dependent Variable: Child is Male

	(1)	(2)
Explanatory Variables:		
Native (0/1)	.003	.024***
	(.37)	(4.19)
Native × Year=1980	.034***	
	(6.76)	
Native × Year=1990	.019***	.006***
	(9.27)	(4.12)
Year=1980	.031***	
	(7.33)	
Year=1990	.034***	.02***
	(6.46)	(18.65)
constant	.489***	.492***
	(85.58)	(125.04)
CONTROLS: Maternal and child age dummies, birth year dummies, martial status dummy		
Number of Observations	5041	3163

Notes: The table presents results of a regression of child sex on whether the mother is an Alaskan Native and census year and interactions. The HBV vaccination program covered Natives beginning in the early 1980s. Regression is limited to children under 4. Data from the IPUMS 1980, 1990 and 2000 5% samples.

[a] t-statistics in parenthesis

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5[a]
Sex Ratio and HBV Vaccination: Taiwan
Dependent Variable: Child is Male

	(1)	(2)
Explanatory Variables:		
Hepatitis B Prevalence (%)	.00056***	.00040*
	(2.76)	(1.90)
Year	.00026***	
	(4.85)	
constant	−.016	.511***
	(−.14)	(144.32)
Cohort FE	YES	YES
Year FE	NO	YES
Number of Observations	96	96
R^2	.49	.68

Notes: Regression is run using panel data on all births in Taiwan by cohort (age cohorts 15-19, 20-24,25-29,30-34) and year(1979-2002). HBV prevalence is calculated for each cohort-year using data on pre- and post-vaccination infection rates (Hsu et al, 1999; Ni et al, 2001; Li et al, 2003).

[a] t-statistics in parenthesis

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6[a]

Cross Country Hepatitis B Prevalence and Share Male at Birth

Dependent Variable: Share of Births that are Male

Explanatory Variables:	(1) All	(2) All	(3) OECD	(4) All	(5) All, Complete
Hep B Prev (%)	.0009***	.0010***	.0014***	.0011***	.0012***
	(4.63)	(4.88)	(4.03)	(3.91)	(9.87)
Log GDP/capita, 1995		−.0020	−.0027	−.0017	−.0016**
		(−1.46)	(−1.32)	(−1.14)	(−2.54)
Total Fertility Rate		−.0008	.0005	−.0002	−.0006
		(−1.04)	(.14)	(−.28)	(−1.52)
Life Expectancy		.0010*	.0003	.0006*	.0004***
		(1.9)	(1.04)	(1.84)	(2.76)
Infant Mortality		.0000	.0000	.0000	.0001*
		(.43)	(−.92)	(.5)	(−1.71)
constant	.512***	.492***	.511***	.486***	.502***
	(717.08)	(29.1)	(23.26)	(26.01)	(63.15)
Region FE	NO	NO	NO	YES	NO
Number of Observations	38	38	18	38	30
R^2	.37	.51	.73	.54	.86

Notes: HBV prevalence is calculated by aggregating published studies; studies used for each country appear in Appendix B. Only countries with more than 2500 people tested for HBV are used in the analysis. Share male at birth is the share of boys born, and data on births are from the United Nations Demographic Yearbook Historical Supplement for all available years from the 1960s through the present. For South Korea and China births in years after 1990 are excluded due to clear increases in use of sex-selective abortion. In Column 5 the data is restricted to country-years with complete birth reporting.

[a] t-statistics in parenthesis

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7[a]

Table 7[a]
Male Births and Hepatitis: IPUMS
Dependent Variable: Birth is Male

	(1)	(2)	(3)	(4)
Explanatory Variables:				
Hepatitis Prev. (%)	.0013***	.0014***	.0013***	.0013***
	(5.37)	(4.77)	(6.36)	(6.66)
Educ. Relative to Cohort			.00089***	.00088***
			(2.78)	(2.73)
Unmarried (0/1)				−.0028
				(−.84)
constant	.509***	.511***	.509***	.511***
	(345.37)	(10.15)	(10.14)	(10.09)
Year of Birth Dummies	YES	YES	YES	YES
Maternal Age Dummies	NO	YES	YES	YES
Number of Observations	210556	210556	194220	194220

Notes: The table presents regressions of child sex on controls, including HBV prevalence in maternal country of origin. Data used are from the IPUMS and are limited to children ages 1-6 born to immigrant mothers. Only countries with HBV prevalence estimated from at least 2500 subjects are included, and only countries with at least 2000 births during the 20th century. Columns 3 and 4 are limited to post-1920 data.
[a] t-statistics in parenthesis
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8[a]

Cross Country Regressions Including Africa

Dependent Variable: Percent Boys Born

	(1)	(2)	(3)
	Africa Only	All	All
Explanatory Variables:			
Hep B Prev (%)	.00077**	.00053***	.00080***
	(2.30)	(2.68)	(3.60)
Log GDP/capita, 1995		−.003**	−.002**
		(−2.46)	(−2.23)
Total Fertility Rate		−.002*	−.001
		(−1.81)	(−.5)
Life Expectancy		.001***	0*
		(3.72)	(1.97)
Infant Mortality		.0000	.0000
		(1.13)	(.58)
constant	.498***	.492***	.505***
	(111.4)	(40.47)	(42.52)
Number of Observations	18	56	56
R^2	.25	.43	.58

Notes: HBV prevalence is calculated by aggregating published studies; studies used for each country appear in Appendix B. Only countries with more than 2500 people tested for HBV are used in the analysis. Share male at birth is the share of boys born, and data on births outside of Africa are from the United Nations Demographic Yearbook Historical Supplement for all available years from the 1960s through the present. For South Korea and China births in years after 1990 are excluded due to clear increases in use of sex-selective abortion. Sex ratio at birth within Africa is calculated from retrospective household survey data (from the Demographic and Health Surveys).

[a] t-statistics in parenthesis

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9
Estimates of Effect of HBV on Sex Ratio at Birth

Estimates from Individual-Level Data		
Mean	1.54	
95% Confidence Interval	(1.29,1.87)	

Estimates from Aggregate Data		
Study	**Estimate**	**Description**
Cross Country Estimates	1.40-1.60	Using regression results for effect of hepatitis on sex ratios as reported in Table 5. The low end estimate is for the overall regression without controls; the high end comes from the OECD. Regressions with overall data and controls yield estimates of around 1.5.
Alaskan Evidence	1.96	Using regression results from the IPUMS, assuming HBV rate of 10% among Alaskan Natives before vaccination

Notes: Individual-level estimates are calculated from an OLS regression of gender on parental HBV status, with study-specific fixed effects. All other estimates calculated as in Equation (4). Discussion of the individual-level evidence is in section 4.2; discussion of the cross country evidence is in section 4.4; discussion of the Alaskan evidence is in section 4.3.

Table 10
Estimates of HBV Prevalence and Predicted Sex Ratio at Birth

Country	Estimated HBV Prevalence	Predicted Sex Ratio at Birth	Range in Empirical Sex Ratio at Birth
China	11.24%	1.103	(1.084-1.14)
India	4.33%	1.069	(1.053-1.11)
Pakistan	2.28%	1.059	(1.025-1.10)
Bangladesh	3.80%	1.067	(1.057-1.076)
Nepal	1.22%	1.054	(0.90-1.14)
West Asia	4.05%	1.068	(1.036-1.091)
Egypt	5.26%	1.074	(1.032-1.112)

Notes: HBV prevalence is calculated from published studies; studies used for each country are in Appendix B. Prevalence reported is for pregnant women in all countries but Pakistan and Nepal. Due to lack of data these are for all healthy individual. Predicted sex ratio at birth is calculated as in Equation (3) in Section 2, assuming a sex ratio of 1.54 for HBV carriers and 1.048 for non-carriers. Range in empirical sex ratio at birth is created from the Demographic Yearbook data (United Nations, 1997) for countries other than India and Nepal; for those countries the range is the range of children's sex ratio across ages as reported in Demographic and Health Surveys from 1990-2000. West Asia includes Turkey and Syria.

Table 11
Re-Estimating the Missing Women

PANEL A: Comparison with Coale (1991)

Country	Coale (1991) Population Sex Ratio	HBV-Adjusted Population Sex Ratio	Actual Population Sex Ratio	Female Population	Coale (1991) Missing	HBV-Adjusted Missing	% Explained
China	1.010	1.052	1.066	548.7	30.42	7.18	76.40%
India	1.020	1.030	1.077	406.3	22.71	18.58	18.17%
Pakistan	1.025	1.025	1.105	40	3.12	3.11	0.28%
Bangladesh	1.025	1.032	1.064	42.2	1.61	1.29	19.64%
Nepal	1.025	1.020	1.050	7.3	0.18	0.21	0.00%
West Asia	1.030	1.039	1.060	55	1.60	1.13	29.54%
Egypt	1.020	1.034	1.047	23.5	0.62	0.29	53.67%
TOTAL					60.26	31.79	47.24%
95% Confidence Interval						(17.2,52.1)	(13.5%,71.5%)

PANEL B: Comparison with Sen (1992)

Country	Sen (1992) Population Sex Ratio	HBV-Adjusted Population Sex Ratio	Actual Population Sex Ratio	Female Population	Sen (1992) Missing	HBV-Adjusted Missing	% Explained
China	0.977	1.052	1.066	548.7	49.98	7.18	85.63%
India	0.977	1.030	1.077	406.3	41.59	18.58	55.32%
Pakistan	0.977	1.025	1.105	40	5.24	3.11	40.60%
Bangladesh	0.977	1.032	1.064	42.2	3.76	1.29	65.67%
Nepal	0.977	1.020	1.050	7.3	0.55	0.21	60.84%
West Asia	0.977	1.039	1.060	55	4.67	1.13	75.84%
Egypt	0.977	1.034	1.047	23.5	1.68	0.29	82.88%
TOTAL					107.47	31.79	70.42%
95% Confidence Interval						(17.2,52.1)	(51.5%,84.0%)

Notes: Coale(1991) and Sen(1992) population sex ratios taken from their papers. HBV-adjusted population sex ratio is calculated using information on sex ratio at birth from Table 9 and model life tables from Coale, Demeny and Vaughan (1983). Actual sex ratio represents population sex ratio in the mid-1980s. The number of missing women (in millions) is calculated as in Equation (1) in Section 2.

Table 12[a]

**HBV and Sex Ratio Across Region: India and China**

Dependent Variable	Share Boys, age 0-6	Share Boys at Birth	Share Boys at Birth
	India	China	China
Explanatory Variables:			
Hepatitis B Prevalence (%)	.0036*	.0010***	.00097**
	(1.88)	(2.50)	(2.31)
Share Agriculture			−.009
			(−.66)
Per Capita Income (yuan)			$-3.55e^{-6}$
			(−1.15)
constant	.505***	.516***	.525***
	(68.42)	(125.82)	(47.18)
Number of Observations	8	60	60

Notes: HBV prevalence for Indian states is from the World Health Organization (2002). For China, data on HBV is from Campbell et al (1990). Data on sex ratio among young children in India is from Dreze and Sen (2003); for China, data on births is from the 1990 Chinese census.

[a] t-statistics in parenthesis

* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1
**Sex Ratio by Age, Missing Women Countries**
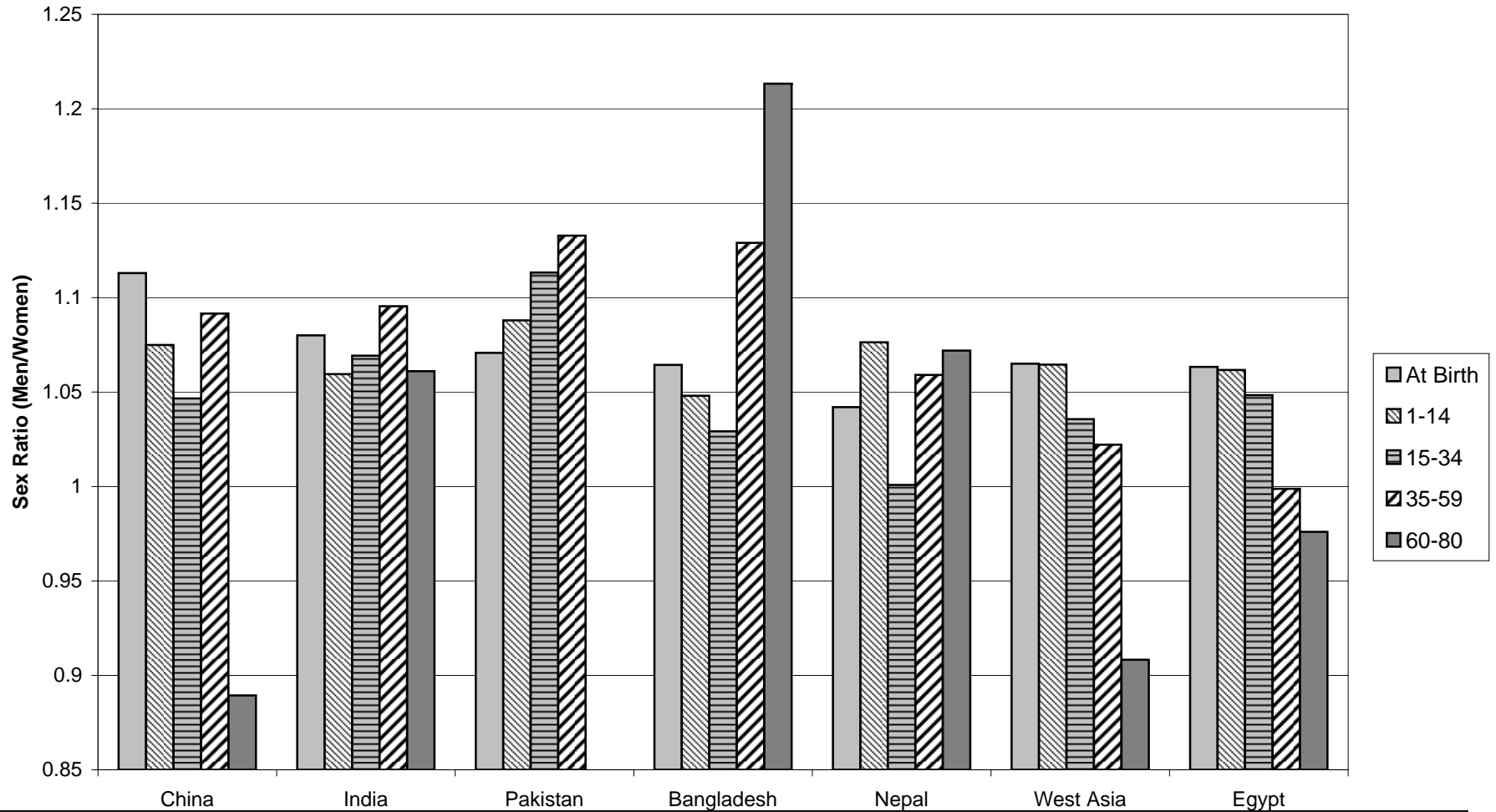
Notes: This figure presents sex ratios by age group for the missing women countries. Data on sex ratios are from the United Nations Demographic Yearbook Historical Supplement (United Nations, 1997) and represent all censuses and birth registration data between 1970 and 1990. Data on sex ratio at birth for India and Nepal is from the Demographic and Health Surveys run between 1990 and 2000 and represents reported sex ratios among children ever born to survey respondents.

Figure 2
Sex Ratio by Age, Indian and Chinese Cohorts
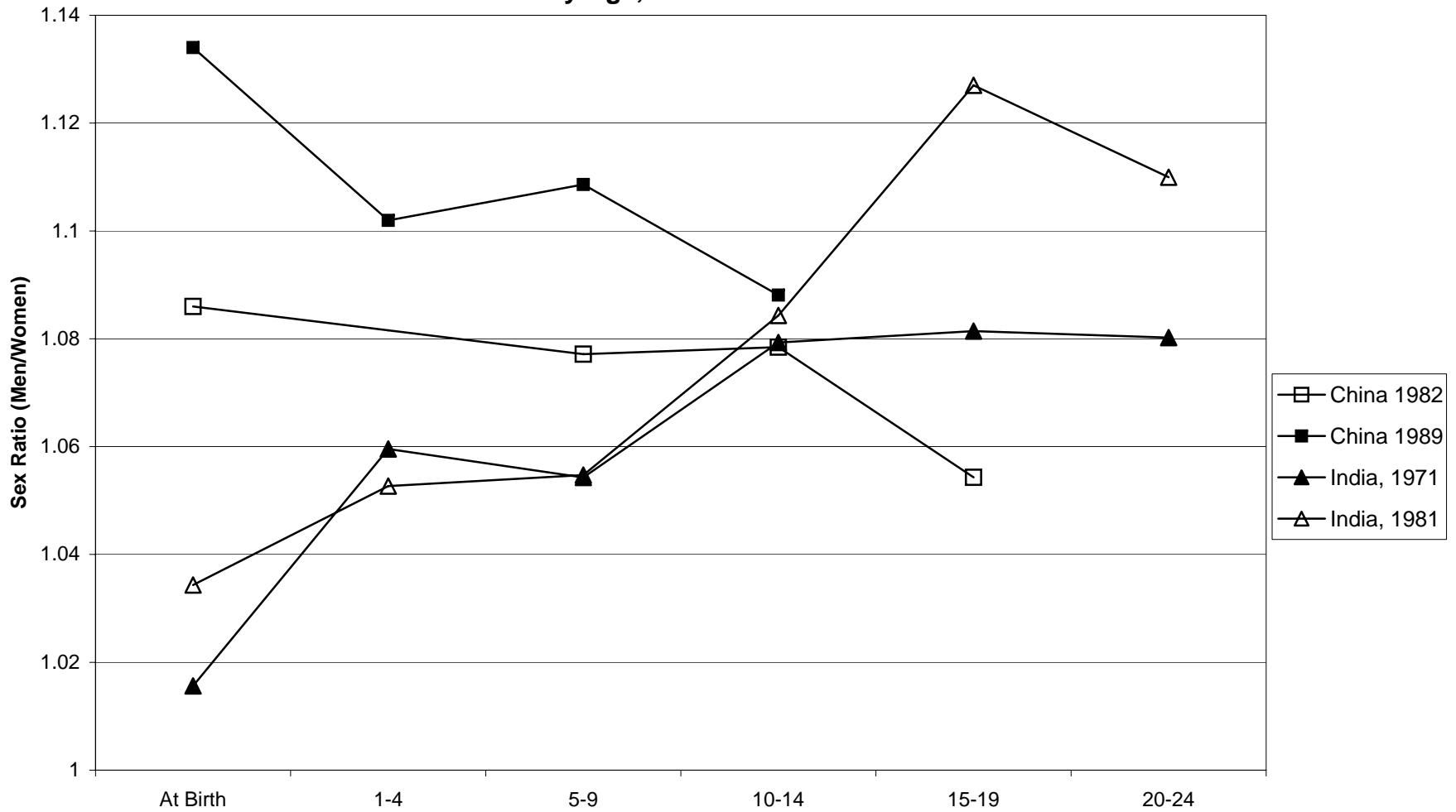
Notes: This figure presents sex ratios by age for specific birth cohorts in India and China. For example, the China 1982 line shows sex ratios at each age for the cohort born in China in 1982. Data for India is from the Demographic Yearbook Historical Supplement (United Nations, 1997) and the Demographic Yearbook (United Nations, 2001). Data for China is from the overall Chinese census and the Chinese 1% census.

Figure 3
**Changes in Sex Ratio at birth in Alaska by Region Type**
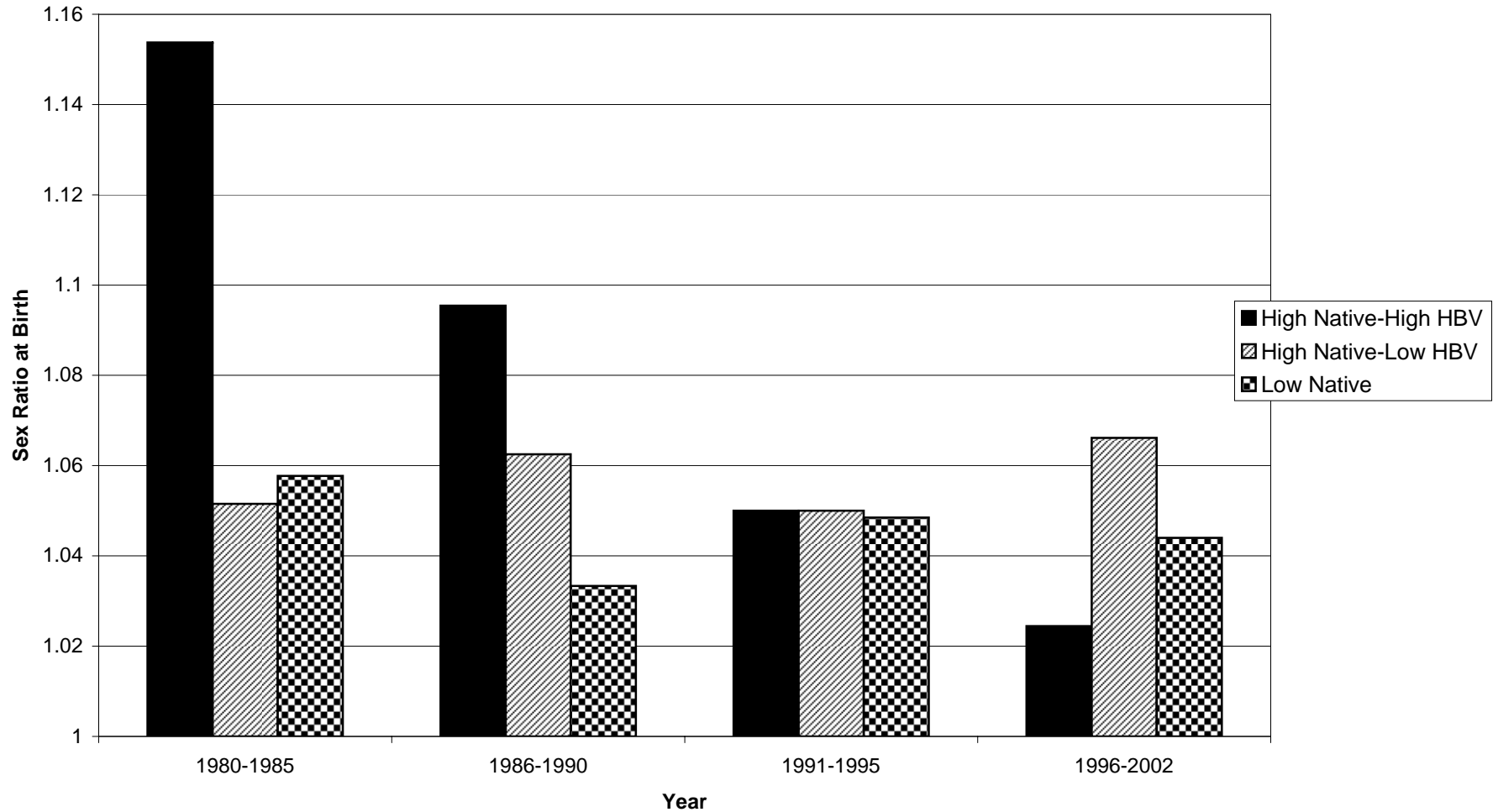
Notes: The figure compares overall sex ratios at birth across three region types: High-native, high-HBV regions (Yukon-Koyukuk, Aleutians East Borough, Lake and Peninsula, Bristol Bay), high-native, low-HBV regions (Prince of Wales, Skagway-Hoonah-Angoon, Wrangall Peterson) and a low-native region (Anchorage). Sex ratio is number of boys born divided by number of girls.

Figure 4
**Share Male at Birth by WHO HBV Prevalence Categories**
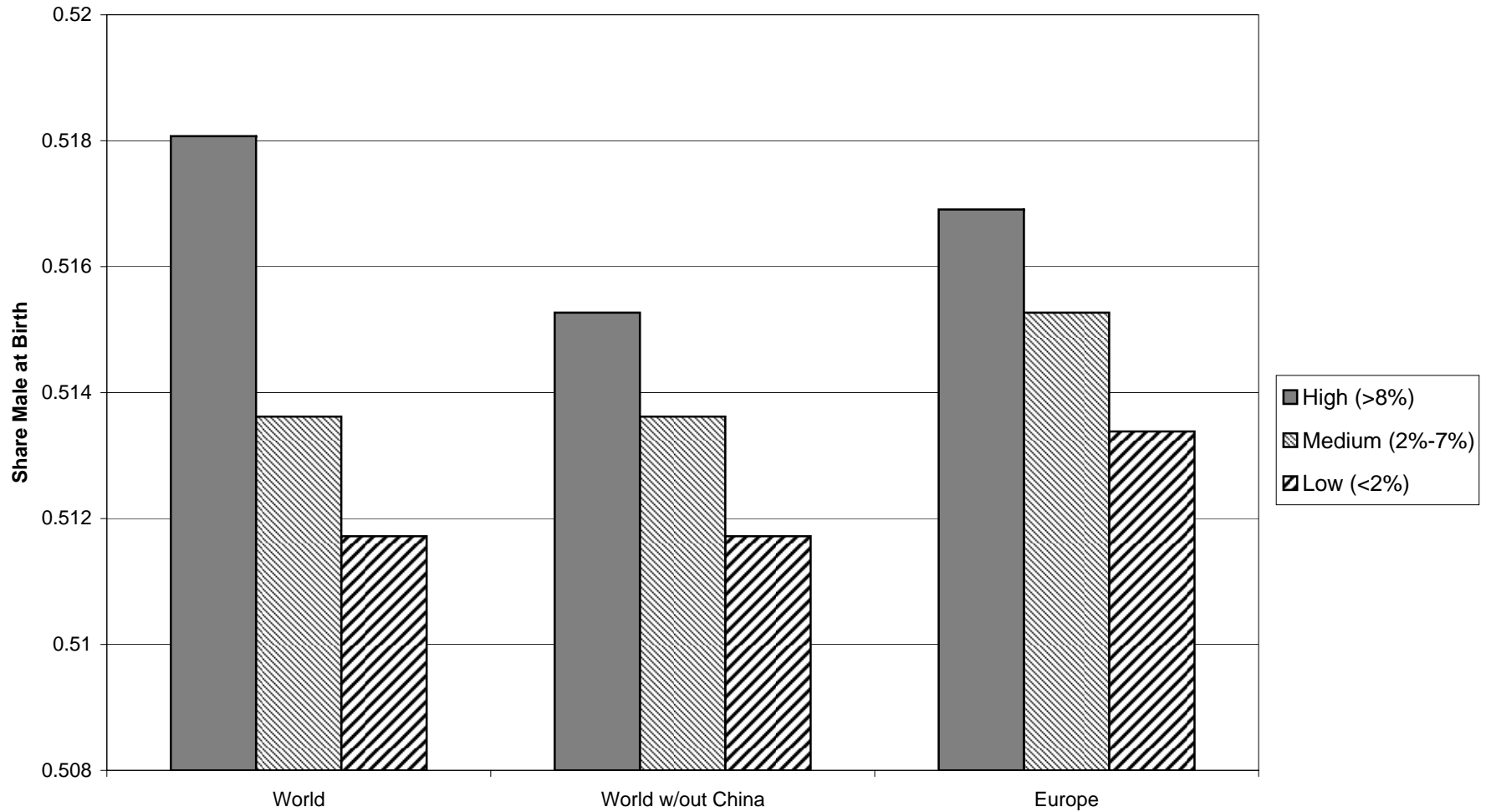
Notes: HBV prevalence information is from the WHO, and countries in each category are in Appendix A. Data on births is from the United Nations Demographic Yearbook Historical Supplement. Share male for each category is weighted by total number of births in each country included in the category.

Figure 5
Share Male at Birth and HBV Prevalence
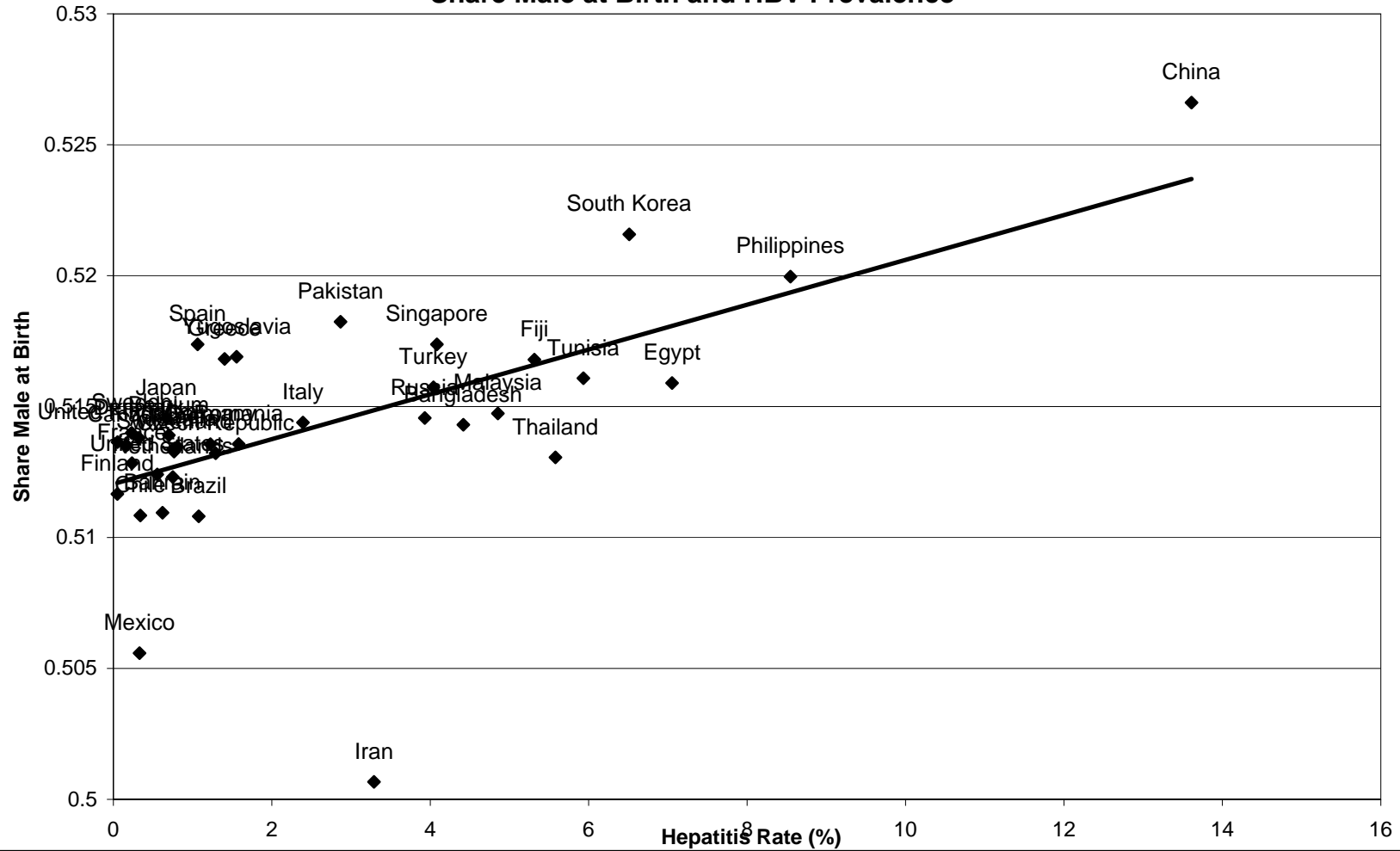
Notes: HBV prevalence is calculated by aggregating published studies; studies for each country appear in Appendix B. Only countries with more than 2,500 people used to calculate HBV prevalence are included.   Share male at birth is the share of babies born who are boys; data on births is from the United Nations Demographic Yearbook for all available years from the 1960s through the present. For South Korea and China years after 1990 are excluded due to clear use of sex-selective abortion technology and large increases in sex ratio at birth.
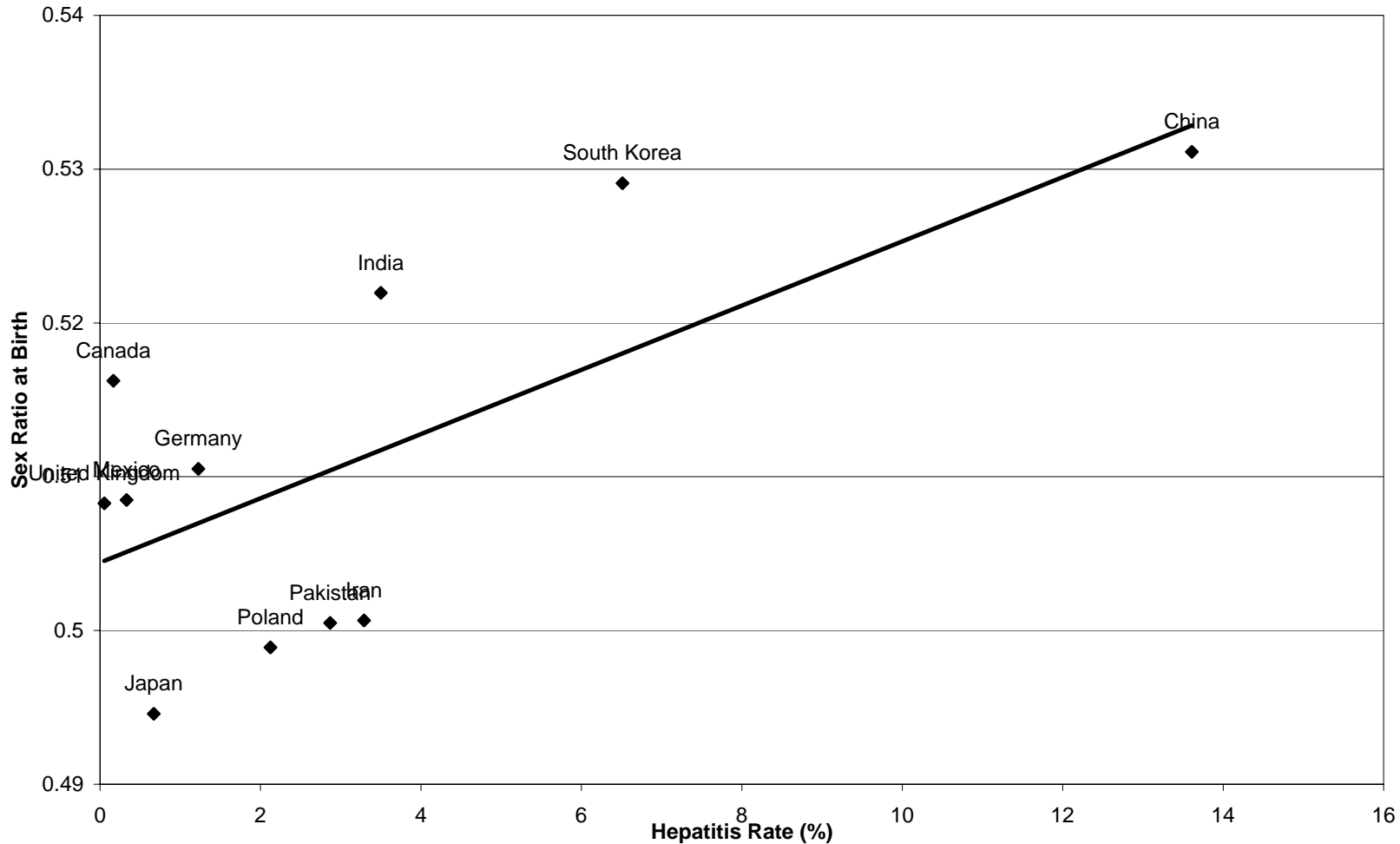
Figure 6
Share Male at Birth and HBV Prevalence, OECD
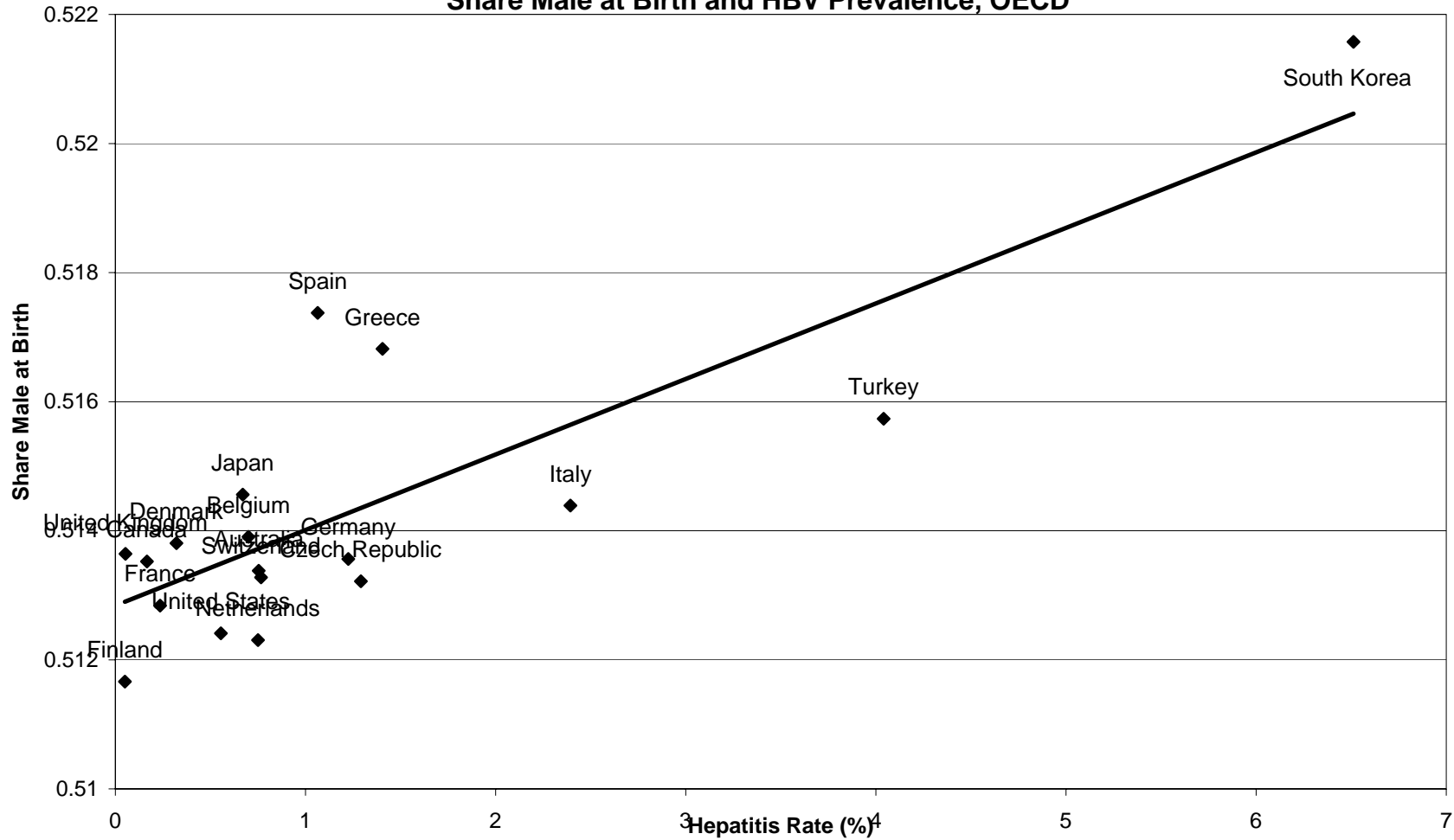
Notes: HBV prevalence is calculated by aggregating published studies; studies for each country appear in Appendix B. Only OECD countries with more than 2,500 people used to calculate HBV prevalence are included.   Data on births is from the United Nations Demographic Yearbook Historical Supplement for all available years from the 1960s through the present. For South Korea data after 1990 is excluded due to concerns about increases from sex-selective abortion.

Figure 7
Share Male and HBV, Data on US Immigrants from the IPUMS

Notes: HBV prevalence is calculated by aggregating published studies; studies for each country appear in Appendix B. Only OECD countries with more than 15,000 people used to calculate HBV prevalence are included. Share male is calculated for each country using children (ages 1-6) of immigrant parents recorded in the IPUMS 5% samples for 1990 and 2000.