

**Consumption, health, gender and poverty**

Anne Case  
Angus Deaton

Research Program in Development Studies  
Princeton University

First version May, 2002, revised June and July, 2002

We are grateful to Shawna Samuel for research assistance and to Andy Mason and Martin Ravallion for helpful comments.

## **Consumption, health, gender and poverty**

**Anne Case and Angus Deaton**

### **ABSTRACT**

Standard methods of poverty measurement assume that an individual is poor if he or she lives in a *family* whose income or consumption lies below an appropriate poverty line. Such methods can provide only limited insight into male and female poverty separately. Nevertheless, there are reasons why household resources are linked to the gender composition of the household; women's earnings are often lower than men, families in some countries control their fertility through differential stopping rules, and women live longer than men. It is also possible to link family expenditure patterns to the gender composition of the household, something we illustrate using data from India and South Africa. Such a procedure provides useful information on who gets what, but cannot tell us how total resources are allocated between males and females. More can be gleaned from data on consumption by individual household members, and for many goods, collecting such information is good survey practice in any case. Even so, we suspect that it will be some time before such information can be used routinely to produce estimates of poverty by gender. A more promising approach is likely to come within a broader definition of poverty that includes health (and possibly education) as well as income. We discuss recent work on collecting self-reported measures of non-fatal health, and argue that such measures are already useful for assessing the relative health status of males and females. The evidence is consistent with non-elderly women generally having poorer health than non-elderly men. We emphasize the importance of simultaneously measuring poverty in multiple dimensions. The different components of wellbeing are correlated, and it is misleading to look at any one in isolation from the others.

## Introduction

This paper is concerned with methods for measuring poverty that allow men and women to be differentially poor. The currently dominant method of measuring poverty, in the World Bank and elsewhere, counts the number of people who live in households whose collective household income (or expenditure) is less than some cutoff. Because these income-based methods rely on *household* measures of resources, they have limited ability to measure differences in poverty between men and women, who typically live together in households. Nevertheless, the literature contains a number of methods that, at least in principle, allow separate poverty counts for men and women using only household-level measures of resources. In Section 1, we discuss some of them, and present some results using data from India and South Africa. All such methods rest on controversial, readily challenged, and not always transparent assumptions. Our conclusion is that, while they often produce results that are interesting in their own right, they do not deliver everything we need. We think that it is unlikely in the foreseeable future that any of the methods will yield a generally useful and acceptable methodology for producing separate counts of the number of males and females in poverty. One possible remedy is to collect more data on *individual* rather than *household* consumption; there is scope for experimentation here, but there are also some real difficulties.

We argue that broader notions of wellbeing, which recognize education and health, are not only theoretically superior, but offer a much more promising route to gendered measures of poverty. Because these measures are gathered for people, not households, direct measurement by sex is immediate. The measurement problems lie elsewhere, particularly for health. Education measures are relatively straightforward; measures of illiteracy, years of education, and

educational attainment are routinely and successfully used to compare males and females. For health, however, summary measures, such as self-reported health status or even counts of health limitations, are far from obviously comparable between men and women, just as they are unlikely to be comparable across income or educational groups, or across countries. In Section 2, we review these arguments, and conclude that the difficulties with self-reported health status can easily be exaggerated. We believe that such measures provide useful comparisons of the health status of men and women, and should be a central component of poverty comparisons by sex.

We also emphasize that health, educational, and income-based approaches to poverty must be implemented simultaneously. Many studies have revealed “gradients” in health and educational status by income, some of which are gendered so that, for example, girls are sometimes less likely to go to school than boys if they live in poor households, but equally likely to go to school in well-off households, Filmer (1999). Poorer people live shorter and less healthy lives than richer people the world over, and excess infant mortality among girls is higher among less educated women. In such circumstances, separately aggregated measures of income, health, and educational poverty, such as those used to measure gender related development in the *Human Development Report*, miss the multiple deprivations associated with the correlations between the different dimensions of poverty. The evidence shows that women disproportionately suffer from poverty in several dimensions simultaneously.

Section 3 concludes and provides recommendations for data collection.

### **1. Income-based poverty measures**

We use the term “income-based measures” to include consumption-based measures, but to

distinguish them from those that use information on health or education. In this section, we review some of the ways in which household measures of resources can inform us about the extent to which males and females are in poverty. Section 1.1 reviews the standard procedure, that of assessing individuals as poor if they live in households whose total resources are below some cut-off, and asks what this can say about poverty by sex. Because resources are not differentiated by sex, differentiation comes about through the interaction between household resources and household demographic composition, and through the extent to which poverty lines are related to the latter. There are several mechanisms through which household resources are systematically related to household composition, including differential stopping rules in fertility behavior, the greater life-expectancy of women, the generally greater earning ability of men, and so on. While poverty lines nowadays rarely differentiate by sex, they are almost always responsive to household size and sometimes to household composition, both of which differ for men and women. It is in this context that the construction and use of equivalence scales can affect the gender composition of poverty.

Section 1.2 turns to methods of using household expenditure data to make inferences about the allocation of resources within the household. Given that there is typically no information in the surveys on the allocation by sex, all such methods rely on assumptions, which tend to be controversial. Even if the assumptions are accepted, implementation of the methods raises a number of econometric issues that are far from being solved. While some of these methods clearly do not work, we argue that a relatively unstructured approach can yield insights and useful information, even if it cannot deliver poverty counts by gender.

Section 1.3 is a discussion of what might be possible if there were more collection of data on

the consumption of individuals, not just households. We argue that, in many countries, household data would be more accurate if more expenditures were collected on this basis, and we review methods from the literature that allow identification of household sharing rules on the basis of such data. We believe that there is good work still to be done along these lines, and we believe that more individual consumption data should be collected. Even so, we are not optimistic that such methods will soon permit routine measurement of male versus female poverty.

### **1.1 Household composition, gender, and poverty**

The most frequently cited poverty measure is the fraction of people living in households whose resources, usually per capita income or per capita total household expenditure, are less than a common poverty line. Table 1, from the 55<sup>th</sup> Round of India's National Sample Survey, collected in 1999–2000, shows some relevant data. The major states of India are shown, together with All India; all data are from the rural sector. The first three columns show mean per capita total household expenditures, for males and females together, for males, and for females. The household per capita expenditure figures are taken directly from the data; we have used the 30-day reporting period, and we have assigned household per capita expenditure to each member of the household. The average over all persons appears in the first column, and the averages over males and females respectively appear in the second and third. The last three columns show the percentages of persons, males, and females, who live in households whose per capita expenditure is below the official (Planning Commission) poverty lines for each state. The “all” column here corresponds (up to rounding error) to the Planning Commission's own estimates of poverty in India. (We have made no attempt to correct the poverty estimates to make them comparable with

earlier surveys; the 55<sup>th</sup> Round estimates are entirely sensible in their own right, even if not comparable with earlier rounds.)

The differences between males and females are modest. Although average per capita expenditure is lower for women than for men (except in Haryana), the average difference is only 10.34 rupees per month, or just over two per cent. This turns into a 1.1 percentage point difference in the headcount poverty rates. Why do these differences exist, and should we attach any significance to them? Given that most households contain both males and females, and given that we are assigning the same per capita expenditure to all members of each household, irrespective of sex, the differences come from the fact that there is a negative relationship across households between the fraction of females and per capita expenditure. This comes in part from the fact that men have higher earnings than women in most cases, so that households that have a higher fraction of females will generally have lower earnings, and thus lower per capita expenditure. This has been much studied in the context of female headship, which often occurs only in the absence of an adult male, and where incomes are likely to be particularly low. But the point holds more generally.

There are also systematic differences in household composition for males and females. For example, in nearly all countries, women have longer life-expectancy than men, so the ratio of women to men increases at higher ages. In rich and middle-income countries, many older women form one-person households which may be at relatively high risk of poverty, depending on the private and public arrangements for social security. In poor countries, such as India, elderly people rarely live alone. The presence of non-earning elders tends to depress per capita income and consumption, and, because there are more elderly women than men, there will be an effect in

the direction of lowering average per capita expenditure among women.

Another demographic mechanism which has been less discussed in the context of poverty measurement is the effect of gender-related fertility rules. In countries where many parents prefer sons to daughters, and where sex-selective abortion is not available, parents can influence the gender composition of their families by continuing to have children until they reach a desired number of sons. If every birth has an equal probability of being a boy (0.513 in India), such behavior cannot affect the gender composition of the population. However, parents who are “unlucky” early in their fertility will continue to have children, so that girls tend to live in large families and have more siblings than do boys. Smaller families have a higher fraction of boys, and the average family size of girls is larger than the average family size of boys. Families with the strongest son preference tend to have largest family sizes, and the highest ratio of girls, by what Shelley Clark (2000) refers to as the “irony of Bernoulli mathematics.” Girls then live in larger than average households, they tend to have a lot of sisters relative to brothers, and they have parents who would have preferred boys. If son-preference is related to low education, more daughters will be born to poorly educated, and relatively low earning parents.

Kazuo Yamaguchi (1989) provides a discussion of the underlying mathematics. For example, Yamaguchi shows that if the probability of having a boy is  $p$  and is constant across families, and if families go on having children until they have  $M$  boys, the expected number of boys in the household is  $M/p$  and the expected number of girls is  $(M+1)/p$  so that, on average, there are  $1/p$  (approximately 2) more girls than boys in such families. Of course, not all families behave in this way, and not all that want to will succeed (for example if they stop being fertile before having  $M$  boys), so that in practice the difference will be smaller.

Table 2 shows household sizes for males and females, as well as for boys and girls, using the same survey as in Table 1. The procedure is the same as before; each person in the household is assigned the household size of the household in which he or she lives, and the results averaged over the different groups. For all males and females, there is little or no difference in household size, only 0.02 persons over rural India as a whole. However, girls aged 0–14 live in households that are 0.16 person larger on average, and the differences are a good deal larger in those states where son-preference is strongest, see Clark (2000) and Murthi, Guio, and Drèze (1995). For example, the household size gap is 0.33 in Gujarat and Haryana, 0.25 in Himachal Pradesh, 0.29 in Madhya Pradesh, 0.32 in Maharashtra, 0.26 in Orissa, 0.25 in Punjab, 0.21 in Rajasthan, and 0.23 in West Bengal. Many of these states are in the North and West of India, where son preference also shows up in excess infant mortality for girls, while the states of the South, Tamil Nadu, Kerala, and Karnataka show much smaller differences, or even a negative difference in Kerala.

What implications do these gender-related differences in household size have for the measurement of poverty in India? Recall that the standard Indian poverty counts, like many others, use *per capita* household expenditure as their measure of resources. Because such a measure ignores both economies of scale and the lower cost of children, it understates the resources available to larger households, particularly larger households with children. Making some sort of adjustment would decrease the poverty rate for larger households relative to smaller ones, and would decrease the measured poverty of girls relative to the measured poverty of boys. Of course, such estimates take no account of intrahousehold allocation, and in particular of the very real possibility that high birth-order girls, born to parents who did not plan for them, are

treated less well than their brothers, Das Gupta (1987).

Whenever the demographic structure of families differs by sex, poverty measures will be affected, not only by how resources vary with household composition, but also by the assumptions we make about how family needs vary with family composition. In India, households with the same level of per capita total expenditure are assumed to be equally well-off, but other countries—such as the US—have more elaborate schemes that tailor poverty lines to household compositions or, alternatively, divide household resources, not by the number of people, but by the number of *adult equivalents*. The number of equivalent adults in each household is typically calculated by assigning a lower weight to children, and by assuming that there are diminishing marginal needs to each additional (weighted) person. For example, one much used formula is to define the number of equivalents  $E$  by the formula

$$E = (A + \alpha K)^\theta \quad (1)$$

where  $A$  is the number of adults in the household,  $K$  is the number of children, and  $\alpha$  and  $\theta$  are parameters that lie between 0 and 1. The parameter  $\alpha$  represents the costs of a child relative to that of an adult, while the parameter  $\theta$  captures the extent of economies of scale. When  $\theta$  is low, economies of scale are large, and *vice versa*.

One useful procedure is to use (1) to examine the robustness of a poverty profile, looking at how the composition of the poor changes as we vary  $\alpha$ ,  $\theta$ , and the poverty line. Deaton and Paxson (2001) use such methods to look at poverty among children and the elderly in a number of countries. As might be expected, the elderly are less likely and children more likely to be poor when  $\alpha$  and  $\theta$  are both large, so that children cost almost as much as adults, and economies of scale are limited. The detailed outcomes depend on the patterns of household composition and

resources in each country, as well as the choice of poverty line. In spite of the variation in household composition by gender that we have seen, it is hard to see that the variation in the parameters of (1) will have much of an effect on the gender composition of poverty provided, of course, that the equivalents are not constructed so as to weight women differently from men. Such gendered scales have been used in the past, defended on the grounds that women are less likely to undertake heavy agricultural work, but they have the effect of essentially *assuming* the gender structure of poverty.

As we have described it here, (1) works by assuming values for the parameters, or at least trying out a range. By contrast, there is a literature dating back to the middle of the 19<sup>th</sup> Century that attempts to estimate equivalence scales by looking at the structure of household budgets (or saving, or time use) for households of different compositions. If such methods could be trusted, they could be used to calculate differential needs, not only by household size and composition, but also by males and females. But all such methods must overcome the formidable hurdle raised by Pollak and Wales (1979) impossibility theorem, that such scales cannot be identified without information on shadow prices of children. Popular empirical methods for measuring child costs, such as those due to Engel and Rothbarth make essentially arbitrary identifying assumptions—households with equal food shares are equally well off, or households with equal expenditures on adult goods are equally well off—that are hard to defend, and that have been far from garnering anything like universal assent. See Deaton (1997, Chapter 4) for description and discussion. The nature of economies of scale seems reasonably located in the existence of public or (semi-public) goods in the household, yet the attempt to model such economies of scale in Deaton and Paxson (1998) runs into currently unresolved empirical paradoxes. Another approach is to ask people

who live in different households how much they need to keep out of poverty, but such methods typically lead to very low estimates of costs, see van Praag and Warnaar (1997), presumably because those who have large families are those who perceive the costs as low and benefits as high. While one can remain hopeful that the last word on this topic has not yet been written, there is currently no procedure in the literature that we could recommend as the basis for estimating reasonable child costs or economies of scale, let alone differential needs of males and females.

## **1.2 Intrahousehold allocation**

In spite of the fact that data come only at the household level, there are a number of procedures for “opening up” the household, and estimating who gets what, the “sharing rule,” see in particular the work by Chiappori and his collaborators, Bourguignon and Chiappori (1992), Browning, Bourguignon, Chiappori, and Lechene (1994). These procedures typically require that, for at least some of the budget, it is possible to see explicitly who gets what, either because the data are collected that way (“men’s clothing” and “women’s clothing”) or because we know that the good is consumed only by one person, or by a subgroup of household members, e.g. “adult” goods. Deaton (1989) used a closely-related idea to estimate the possible discrimination against girls by asking whether the reduction in adult goods associated with the birth of a female child was less than the amount of the reduction following the birth of a male child. The data could then be taken as showing the extent to which adults make more room in the budget for boys than for girls, something that would be consistent with other evidence of discrimination against girls, most notably excess infant mortality.

This methodology has by now been widely applied, including to a number of countries such

as Pakistan and Bangladesh where such discrimination is to be expected. Yet the results have not produced any evidence of discrimination; even in those cases where the reduction in expenditures on adult goods is quite precisely estimated, there appears to be no difference between boys and girls, see for example Deaton (1997, p. 240). It is not clear whether there really is no discrimination, or whether, for some reason that is unclear, the method simply does not work. In order to try to gain some further insight into the issue, we have applied the method to the Indian NSS data from the 55<sup>th</sup> Round.

The details are as follows. We specify a model of the form:

$$w_{ih} = \alpha_i + \beta_i \ln \frac{x_h}{n_h} + \xi_i \ln n_h + \sum_{j=i}^{J-1} \gamma_{ij} \frac{n_{jh}}{n_h} + \theta'_i z_h + u_{ih} \quad (2)$$

where  $w_{ih}$  is the share of the budget devoted to good  $i$  by household  $h$ ,  $x_h$  is total household expenditure,  $n_h$  is household size (so that  $x_h/n_h$  is per capita expenditure),  $n_{jh}$  is the number of people in each of  $J$  age and sex groups (note that the last group needs to be dropped from the regression) and  $z_h$  is a vector of other relevant household characteristics. The results for three adult goods, tobacco, intoxicants (alcohol, toddy, ganja), and *pan*, together with their sum, are shown in Table 3.

These estimates provide even less insight into gender discrimination in consumption than did previous results. If we start with “all adult goods” combined, the procedure fails at the first step; the presence of children, except possibly of boys aged 5 to 14, does not seem to exert a well-defined negative effect on expenditures on adult goods. Although the presence of adult males exerts a large positive effect on consumption, and the presence of adult females a smaller but still significant effect, the presence of children has about the same effect as the presence of an elderly

woman. Although the coefficient on boys aged 5 to 14 (the *only* significant coefficient for children) is indeed more negative ( $-0.377$ ) than the corresponding coefficient on girls ( $-0.195$ ), they are not significantly different from one another.

The tobacco regression appears to say that all children enhance consumption relative to the elderly female omitted category, perhaps because the presence of the last acts so as to decrease consumption. Conditional on this odd pattern, it is once again true that there is more tobacco consumption associated with girls than with boys, so the differences in the coefficients are in the direction that discrimination would predict, although not significantly so. The results for *pan* show negative effects of the presence of children, but also for *all* age and sex groups. According to this, old women are the great consumers of *pan*. Here the reductions in consumption associated with boys are very similar to those associated with girls. Finally, the results for intoxicants look like those for tobacco, with all groups consuming more than older women. And once again, although the coefficient for boys is here smaller than that for girls, there are no significant differences between them. Indeed, in none of these regressions are the male and female coefficients for children significantly different, either in single pairs, or all three pairs taken together. We have run these regressions state by state (there are more than 70,000 observations in Table 3), but with similarly discouraging results. As in previous work with this method, there is no evidence for discrimination in consumption.

A number of authors, including Chesher (1997), Bidani and Ravallion (1997), and Deaton and Paxson (2000), have proposed schemes for breaking up an observed aggregate into its components, either for a household or national aggregate. In principle, this offers a straightforward way of disaggregating household expenditure, an approach that has recently been

implemented by Mason, Montenegro and Khandker (1999). According to this, we observe total household expenditure, or expenditure on a specific commodity, which is envisioned as the sum of expenditures by each of the household members, where each person of a given age and sex group receives the same amount. As we shall see, this interpretation cannot be taken seriously, but it provides a motivation for writing

$$x_{ih} = \sum_{j=1}^J \varphi_{ij} n_{jh} + \varepsilon_{ih} \quad (3)$$

where  $x_{ih}$  is expenditure on good  $i$  by household  $h$ , and, as in equation (2), the  $n_{jh}$  are the numbers of people in household  $h$  in each of  $J$  age and sex groups. Perhaps the best way to think about equation (3) is to imagine a cross-tabulation in which each cell corresponds to an age and sex composition, such as one adult female, one adult male, and two young boys, and in which average expenditure for the good is recorded in each cell. Equation (3) is then a simple regression that attempts to fit the mean in each cell by a linear, no constant, function of the numbers of people in each class.

Before further discussing the interpretation of (3), it is useful to look at some results. We use the same age groups as Mason et al, 0–5, 6–15, 16–35, 36–55, and over 55, and we fit (3) to the 10 resulting age-sex classes. We illustrate the estimated  $\varphi$  coefficients as a series of bar charts. The data are the NSS data for rural India in 1999–2000, seen above, as well as those from Black and Coloured households from the 1995 Household Income and Expenditure Survey of South Africa. For the commodities, we examine goods that are either adult or child specific, such as alcohol and tobacco, or adult and child footwear and clothing, as well as “merit” goods, such as educational and health related expenditures. We also examine some broad aggregates such as

food, or total expenditures.

Starting with India, Figure 1 shows total consumption expenditure, where the only major difference by sex is in the 36 to 54 age group, where the yellow male bar on the left is higher than the female bar on the right. Figures 2 and 3 show the corresponding figures for each of the major states of India. The all India pattern reappears in some, but not in all, though there is no obvious pattern in which the favoring of men is greater in the states where women typically do worse, for example, in infant mortality. Figure 4 for South Africa shows a pattern for total expenditure that is similar to that for India, though with a somewhat greater relative advantage for middle-aged males. Figure 4 also shows the South African food data, with once again an apparent advantage for males in the 36-54 year old group, though not at other ages. Mason et al find similar results for food expenditures in Bangladesh. One might use such figures to argue that men, especially middle aged men, get more than women. However, this does not necessarily follow. Men earn more than women, so that such results are consistent with a pattern in which everyone in each household consumes exactly the same amount in total, irrespective of gender, but people in households with more men consume more than people in households with fewer men.

Figure 5 shows tobacco consumption in India, and here the picture is very different. Expenditures on tobacco are clearly associated with the presence of adult males though, as in Table 3, there are surprising positive effects associated with children, especially young girls. Figure 6 repeats the tobacco graph together with the graphs for pan and for intoxicants. Pan is much more equitably associated across the genders, while intoxicants, like tobacco, are associated with men and not women. For both pan and intoxicants, adult goods are indeed

associate with adults! Figure 7 shows the corresponding results for tobacco and alcohol in South Africa. Almost all of these expenditures are associated with adult males, and none with children or with females. Expenditure on personal items, by contrast, is also directed towards adults, but is more equitably distributed by sex, with young women favored over young men.

We turn to educational expenditures in Figures 8 (India) and Figure 9 (South Africa). In India, and in spite of differential enrolment rates by girls and boys, these graphs show no gender-related differences. (But note that gender differences in school attendance have much diminished in India in recent years.) The youngest children, who neither go to school, nor contribute earnings, lower educational expenditures, which is how we interpret the negative values in the two figures. Older children, who go to school, generate educational expenditures, as do adults, who earn the money to pay for them. Figure 9 shows total educational expenditure for South African households, and shows a similar pattern to Figure 8 for India, although there appears to be a gender specific positive effect associated with women aged 36 to 55, an effect that also appears in the subcategories day care and school fees. Perhaps women are more interested in the education of their children and grandchildren than are men, but the interest is normally attributed to female pensioners, over the age of 55, who also have the money to help pay. No such association appears in Figure 9.

Figure 10 shows the results for medical expenditures in India, and shows no pattern of gender differences. The state results in Figures 11 and 12 are more variable, but we have not been able to discern any consistent pattern of discrimination against girls. While girls are associated with less medical expenditures in Madhya Pradesh and Maharashtra, and with more such expenditures in Kerala, there are no anti-girl patterns in Gujarat, Rajasthan, or Uttar Pradesh, where we might

otherwise expect to see them. The South African data in Figure 13 do indeed show a pro-male pattern, but those who are favored are middle-aged men.

The final two graphs, Figures 14 and 15, show footwear and clothing in South Africa, split up by children's and adult clothing, which are separately recorded in the survey. These graphs show no discrimination among children, but as was the case for medical care, middle-aged men are favored over middle-aged women in both adult clothing and footwear.

What can we infer from these graphs? They clearly tell us something about gender patterns in consumption. For example, it is men and not women who smoke and drink, a finding that is credible and supported by a great deal of other evidence. Can we then push them far enough to tell us whether, and to what extent, women get less than men, and boys get more than girls? Unfortunately, the answer is clearly no. As we have seen, these results are affected not only by "needs" but also by the ability to pay for them, so that when we project expenditures on to household composition, we are looking at the effects of needs and earnings simultaneously, so that we cannot tell whether men appear to be favored because they are consuming more, or because they are earning more.

To see what is going on, sum equation (3) over all expenditures to get

$$x_h = \sum_{j=1}^J \varphi_j n_{jh} + \varepsilon_h \quad (4)$$

where  $x_h$  is total household expenditure on all goods, and  $\varphi_j$  is the sum over  $i$  of the individual  $\varphi_{ij}$ . The bar graphs for equation (4) appear in Figure 1 for India and Figure 4 for South Africa, and they show relatively little difference by gender. Suppose, for the moment, that they showed none. Then the comparison of (3) and (4) shows that, if we were to draw graphs for all

commodities, and then add up, we would lose all gender differences in the aggregation, and we would get no overall difference between men and women. But this is not quite what happens. Men tend to earn more than women, and the  $\varphi_j$  for middle-aged men may be larger than that for middle-aged women. But in this more general case, the sum over all goods will not be the same for men as for women, but it will add up to whatever is the difference in their patterns of total consumption, or total household income which, in the absence of saving, is the same thing. The fundamental issue here is that, apart from saving, the individual expenditures must add up to income, and that identity holds no matter what is the composition of the household. So changing the gender composition of the household can only *reallocate* the total budget, not change it, and we have no way of telling whether these reallocations are the consequence of discrimination or, more benignly, of different preferences by gender. But the only way in which the sum of the expenditures can be lower for women is if income (or more precisely income less saving) is lower for women. Indeed, this is exactly the familiar rock that has sunk almost all attempts to measure child costs or gender differences in household budgets. Within the same budget, differences in one direction in one expenditure must be offset by equal and opposite differences somewhere else.

### **1.3 Using individual in place of household data**

Most household expenditure surveys collect data on *household* expenditures or consumption so that, for example, we know how much rice the members of the household consumed in the last month, or how much its members spent on transportation. Such data are often collected from a single respondent, usually the person deemed to be most knowledgeable about such things, who

reports on behalf of all household members. An alternative methodology, sometimes employed, is to ask each adult member of the household about his or her own spending so that, in principle, we can observe, not only total household consumption, but also the consumption of each household member. Such data hold out the promise of constructing individual welfare measures and, in particular, measures by gender. They also provide often invaluable information on the individual's consumption of specific commodities that we are interested in, such as educational or medical expenditures, or the consumption of tobacco.

Of course, many consumption items cannot be allocated to individuals, even in principle. Household public goods, most notably housing, are shared and there is no non-arbitrary method of allocating them to individuals. Private goods are often purchased by someone other than the final consumer, so that a person using the good may not be well-informed about its purchase, nor the purchaser about its use. Food consumption, which in poor households may be more than two-thirds of the budget, is an example of a (mostly) private good—what one person eats another cannot—but one where the observation of individual intakes is notoriously difficult, invasive, and expensive, although certainly not impossible. Yet there are other goods, such as cigarettes, where it is relatively easy to find out who consumes how much. The goods for which individual consumption expenditures are most easily collected are those items where individuals make their own purchases outside the home, such as transportation, personal services, meals taken away from home, and minor purchases made with “walking around money.” Indeed, our own experience is that, in some countries, such items can be accurately recorded only when each member of the household is separately questioned; there is no single household member who knows about all purchases. In the surveys that we have developed to deal with situations where

private knowledge is important, we have a “household” questionnaire in which most household consumption is collected, but also a series of “individual” questionnaires where each adult is asked to report a subset of expenditures.

The promise of individual consumption data lies not only in increased accuracy of total household consumption, but also in the information they contain on allocation within the household. If all goods were private, and were attributable to specific individuals, we could immediately derive total consumption for each household member, and no special techniques would be required. But when we only have a subset of private goods, matters are not straightforward. In particular, if individual  $A$  gets more of some good than individual  $B$ , we have no way of knowing whether the difference is simply a matter of taste with individual  $B$  getting more of something else, possibly even of the shared, public goods, or whether individual  $B$  is generally deprived, getting less than his or her share of household resources overall.

As noted in the previous section, Chiappori and his collaborators, Chiappori (1988, 1992) Bourguignon and Chiappori (1992) have developed a method that allows the recovery of the “sharing rule” given data on the amounts of private consumption. For example, if there are two people  $A$  and  $B$  in the household, and there are three kinds of income,  $y^A$ ,  $y^B$ , and  $y^O$  for  $A$ ’s income,  $B$ ’s income, and “other” income, Chiappori writes the sharing rule in terms of the amount that  $A$  gets as

$$\theta^A (y^A, y^B, y^O). \quad (5)$$

The amount received by  $B$  is total income less the amount received by  $A$ . (Note that “other” income is not simply “unclassifiable” income, but a third category with a stable effect in the sharing rule.) Chiappori then shows that, provided we have data on at least one good whose

consumption by  $A$  and  $B$  is separately available, or two goods that are known to be consumed by only one person, variation in the three types of incomes identifies all three derivatives of the sharing rule. Additional assumptions would identify the rule itself. For example, if (5) were linearly homogeneous, so that a doubling of all incomes doubles both partners' allocations, the sharing rule is the sum of the derivatives multiplied by the respective incomes.

Another way of getting at sharing rules is to see what happens as household composition changes, an idea that goes back at least as far as Rothbarth (1941), and was formalized as the concept of demographic separability in Deaton, Ruiz-Castillo and Thomas (1989). Here we illustrate with the simplest example. Suppose that there are only adults in the household, and that there are  $f$  women and  $m$  men. Suppose too that each person gets "shares" in income, that the shares are the same for all women and for all men, but may differ by sex. Suppose that each woman gets  $\alpha^f$  shares and each man  $\alpha^m$  shares. In consequence, each woman's part of total household income  $y$  is

$$\frac{\alpha^f y}{\alpha^f f + \alpha^m m} = \frac{y}{f + m\rho} \quad (6)$$

where  $\rho$  is the ratio of  $\alpha^m$  to  $\alpha^f$ , and thus captures the gender bias in the household allocation.

Suppose then that we have, as in Chiappori's set up, one or more goods whose consumption by males and by females we can separately observe. For such a good, labeled  $i$ , write the household demand by women as

$$q_i^f = f g_i \left( \frac{y}{f + m\rho}, p, z \right) \quad (7)$$

where  $g_i$  is the demand function for each woman,  $p$  is a vector of commodity prices, and  $z$  other

variables affecting demand. If we add an extra man to the household and, if for the purposes of exposition, we are allowed to have a continuous number of men, we can compare the effect with the effect of an additional unit of income, then we get

$$\frac{\partial q_i^f / \partial m}{\partial q_i^f / \partial y} = \frac{-\rho y}{f + m\rho}. \quad (8)$$

Equation (8) is essentially the “income equivalent ratio” defined in Deaton (1989); it is the amount of additional income that would have the same effect on women’s consumption as adding a man to the household. Note that the right hand side of (8) is independent of  $i$ , so that there are potentially testable restrictions when there is more than one good whose purchases by women we observe.

The corresponding income equivalent ratio for the effect of a woman on men’s goods is

$$\frac{\partial q_i^m / \partial f}{\partial q_i^m / \partial y} = \frac{-y}{f + m\rho}. \quad (9)$$

The ratio of the two income equivalent ratios, of (8) to (9) is the distributional parameter  $\rho$ , which is therefore identified provided we can identify the derivatives with respect to men, women, and income.

We suspect that these methods are worth serious investigation in cases where there is a reasonable number of goods whose consumption can be attributed to men and women separately. Applications of the Chiappori method to date have been hampered by this lack, and have relied heavily on the identification provided by a few goods, such as men’s and women’s clothing. But, even with a richer menu of assignable goods, it may be hard to obtain convincing results. There are (related) theoretical and econometric issues. On the latter, the Chiappori method requires the

ability to exogenously vary the three kinds of income, either directly, or through instruments. The method outlined immediately above faces similar (and perhaps even more difficult) challenges with varying household composition. When an additional person joins the household, all sort of household arrangements are likely to change, including work patterns, child care arrangements, the allocation of time, as well as saving and wealth, none of which is adequately captured in a model that assumes that, conditional on income, an extra person is simply an additional consumer whose needs are met along with those of everyone else of the same sex. Similarly, households with different balances of men's and women's incomes will also have different patterns of labor force participation and work-related expenditures, so that additional "female" income may have a much more profound effect on expenditures than its role in the sharing rule would suggest. None of these problems is insuperable, but much remains to be done, even if there were much more data available than is currently the case.

## **2. Measuring health poverty**

If we cannot make an adequate assessment of poverty based on consumption, then broader approaches may be more successful. In particular, measures of health and education are collected at the *individual* level, not for households, so that they have an enormous immediate advantage over measures based on consumption, income, or wealth. We shall say little about education here, if only because the measurement problems do not seem to be particular severe. Many surveys routinely collect data on literacy, on years of schooling, and on educational achievement, and there is no difficulty in comparing these measures across males and females. The measurement of health is much more difficult, and it is the issue that we discuss in this section.

The standard measure of population health is life expectancy. At any given moment, someone is either alive or dead—states that are generally straightforward to measure and tell apart—so that, by looking at mortality rates by age, we can calculate how long a person is expected to live. Good use has been made of such measures to supplement income-based measures of wellbeing. Even so, life expectancy has a number of limitations as a measure of welfare. It can only be measured at the population or sub-population level, because the estimation of mortality rates requires a large sample of people, and is not useful at the individual level, at least if we are not prepared to wait until the person’s life is over. In order to measure the health of individuals when they are alive, sometimes called “non-fatal health,” we need a measure of health status. Even if life-expectancy could supply this, it is not conceptually what we want. Women typically have longer life-expectancy than men though, remarkably, the *Human Development Report* refuses to recognize the advantage by scaling their Gender Development Index to eliminate it. Yet longer life-expectancy says very little about the burden of non-fatal disease through life and, as we shall see, the evidence suggests that these burdens are higher for women.

A good measure of health status among the general population has been something of a holy grail among researchers, and in Section 2.1, we briefly review the various alternatives that are available. We conclude that the conceptually appropriate measure is some version of self-reported overall (or global) health status, where people report their health on an ordinal scale. We review the various difficulties with such a measure, particularly the argument that it is not comparable across different groups so that we cannot use it to compare, for example, Africans with Indians, rich people with poor people, or men with women. We argue that such difficulties are not severe enough to prevent their serious consideration in the context of poverty

measurement. Section 2.2 presents some evidence on self-reported health status from the Langeberg survey in South Africa, with some comparisons from the United States, as well as from other related studies in the literature. These results support the view that the differences in self-reported health status between men and women are a real component of their differential wellbeing. Section 2.3 discusses some recent work from the World Health Organization on methods of improving the comparability of self-reported health across different groups, and we argue that the WHO surveys, with appropriate supplementation for information on education and on consumption, would be good vehicles for Bank measures of poverty that would not only improve and broaden our understanding, but that would allow poverty to be separately monitored for men and for women.

## **2.1 Measures of health status among the living**

Survey researchers have many ways of collecting information about health status. Self-reported measures can usefully be separated from those made by trained medical personnel, although limited examinations, either in the household or in a medical facility, are now often included as part of a survey effort. Increasingly too, investigators have the ability to merge survey data with medical records. Such records and physical examinations are clearly the appropriate gold standards for conditions that are susceptible to direct measurement, for example, anemia, hypertension, helminthic infestation, or TB, but often tell us relatively little about the extent to which less than perfect health inhibits the way that people function in their lives. People with a clean bill of health after a physical examination may nevertheless be compromised in their ability to lead a full life, for example by severe depression, while some of those who show up with well-

defined conditions may suffer very little functional limitation. The medical construction of health, although important as an input into a healthcare strategy, is much less relevant for measuring wellbeing.

For these and other reasons (including the relatively high cost of examinations by medically trained investigators), many surveys ask the respondent to provide information about their own health. This often takes the form of asking people to report medical conditions that have been previously diagnosed by a nurse or a physician. Such procedures can only work in places where people have regular contact with a healthcare system, but such is not the case in many low-income settings. Such reports may also be quite different from medical records. Although the latter are sometimes incomplete or incorrect, there is almost certainly misreporting by individuals, who may forget what they were told, or may never have been provided with the information in their records.

People can also be asked directly about aspects of their health. Height and weight are relatively straightforward examples, although these can also be directly ascertained by enumerators, and see Thomas and Frankenberg (2000) for documentation of systematic self-reporting biases in the United States. Data on height and weight allow construction of standardized measures of “stunting” and “wasting,” that have been invaluable in assessing the health status of children. Appropriately standardized, these measures are thought to be comparable across boys and girls, as well as across countries. Because height is a measure of early-life nutrition, it retains its relevance for adult health, as does weight, because both high and low weight relative to height are risk factors for morbidity and mortality in adults. Yet neither height nor weight for height within the large normal range provide much of an indication of

functioning among adults.

Many health-related surveys collect data on “activities of daily living” (ADLs) or “instrumental activities of daily living” (IADLs). The ADL questions relate to activities that are central to people’s daily functioning, such as the ability to feed, clothe, bathe, or toilet oneself, or to rise unaided from a chair. IADLs, as their name suggests, relate to activities that are instrumental to the more basic functions. Examples might be the ability to walk for a few kilometers, carry a heavy load, fetch water, catch a bus, manage money, or sweep a floor. Respondents are usually asked whether they have any difficulties or limitations with ADLs and IADLs, and a crude measure of health status constructed, for example, by counting the number of reported limitations. The inability to function, as measured through ADL limitations, is closer to what we are looking for in a measure of health poverty. Such measurements are a direct response to the question of the extent to which poor health burdens people’s lives, and would surely be a part of any suitable health measure. However, in most surveys, ADL limitations are relatively rare, except among the elderly, even among people whose health is clearly far from perfect. Poor people, even those who are visibly malnourished, have an astonishing ability to perform even the most grueling daily tasks. More generally, the willingness to let poor health compromise one’s ADLs may respond to other circumstances of life. For example, a headache may only be debilitating if one has the luxury to make it so.

Limitations to IADLs are more common, but their degree of “instrumentality” may vary from person to person. For example, a rich woman may be unable to carry a heavy load—and indeed may express gratitude that she would never be expected to do so—while her poor cousin’s life might be severely compromised by her inability to do so. The link between IADLs and health

poverty is less direct than the link between ADLs and health poverty.

Older health surveys tend to ask people whether they have been “ill” over some response period, say the last thirty days, often with some measure of severity or chronicity, such as whether they were forced to spend time in bed, to miss work, or whether the condition is a recurrent one. But such questions have been much discredited in recent years, again because whether people perceive themselves as ill at all, let alone whether this illness is sufficient to interfere with work, is likely to vary systematically according to their circumstances. One problem is adaptation, that after a while people get used to whatever state they are in, so that the reported severity of an illness will diminish with the time it is experienced. This can lead to the paradoxical effect that people who are chronically sick may report themselves in better health than those who suffer the same disease but less frequently. Bed days, or time spent away from work, may also be a luxury that only those who are better-off can afford, so that patterns of reported morbidity may be more severe among the rich than the poor, even when the opposite is true. Chen and Murray (1989) documented this phenomenon in India, where reported morbidity is much higher in Kerala than in Bihar, even though life-expectancy, as well as a host of anecdotal evidence, indicates precisely the opposite, see also Sen (2002). It is less clear that ADLs or IADLs, which relate to more objective and well-defined situations, are subject to biases that are as severe. Even so, “objective” conditions also get misreported, not only when Americans optimistically “shade” self-reported height and weight, but also in Ghana where people “forget” that they are missing fingers and toes, Belcher et al (1976).

In recent years, much use has been made of self-reports of “global” health status. Respondents are asked to rate their overall health status on a scale, known as a Likert scale,

usually of five points, with descriptions corresponding to something like “excellent, very good, good, fair, and poor.” At least in part, this question owes its popularity to the low cost of including it in a survey, and the ease and speed with which respondents are able to answer it. The use of the measure has also been spurred by studies that show its ability to predict mortality for several years after the report, even conditional on a direct examination by a physician, see Idler and Benyamini (1997) for a review. Yet for our purposes here, predicting mortality is not the main goal. We are much more interested in whether such reports of global health are useful as measures of what they purport to be, which is the individual’s actual health status, as perceived by themselves. The problem is not in the definition, which is exactly what we want, but rather in the question of whether such reports can be treated seriously, whether they are comparable between men and women, or whether, like the illness measures, they are so affected by people’s position (including their health itself) as to make them useless or misleading.

## **2.2 Self-reported health status: some evidence**

Perhaps the greatest difficulty with assessing the validity of self-reported health status (SRHS) is the lack of a gold standard for comparison. While SRHS can be used to predict mortality, or to compare with the results of medical examinations, neither positive nor negative results would be conclusive, because it is conceptually different from either. So there is little alternative but to examine such reports, and to see how they behave in practice. A number of papers in the literature have provided relevant evidence, and we present more here, relying primarily on a recent survey from the Langeberg health district in South Africa, with some comparative figures from the United States.

We begin with the United States, where we have a large number of observations, which serves as a useful baseline. Figure 16 shows ten years of data from the National Health Interview Survey (NHIS) from survey years 1986 through 1995. SRHS is reported on a scale of 1 (“excellent”) to 5 (“poor”) so that, in this case, more is worse. The figure shows the average reported values using the “natural” scale, 1 through 5, for men and women separately by two year age groups from ages 18 through 90. Nearly 800,000 people are represented in the graph.

Up until age 65, women have worse average self-reported health status than men. At age 20, the difference is about 0.2, and this narrows with age until age 65, after which there is no consistent difference by sex although, if anything, women have the advantage. Average SRHS becomes steadily worse with age so that, if there is adaptation by age, with the question being effectively interpreted as “relative to other people your age,” it is not enough to offset the natural deterioration of health with age. Using the 1 through 5 scale, as here, the deterioration of health with age is slow until age 30, when few people report anything other than “excellent” health, and then speeds up in middle-age, slowing down again after age 65. There are many possible explanations for such an effect: perhaps adaptation to declining health is better after age 65, perhaps mortality selects out more rapidly those in worse health, or perhaps Medicare helps ameliorate the effects of aging.

Figure 17 shows comparable data from the Langeberg health district in the Western Cape province of South Africa. These data were collected in 1999 by the South African Labour and Development Research Unit of the University of Cape Town. The figure is drawn for the black and coloured sample only. Because the sample is so much smaller than for the US, only 519 people, we show the results as a bar chart for ten year age groups. Overall, these South Africans

report themselves as being in worse health than Americans, with an average SRHS of 2.82 for blacks and 2.31 for coloureds, compared with 2.23 for Americans. As for the US, women report worse health than men. Once again, the difference vanishes for the group aged 58 to 67, and for (the small sample of) those over 67, women's reported health is better than men's reported health. One of us, Case (2002), has speculated elsewhere that the generous social pension in South Africa might well be related to health among the elderly, and especially to the health of women, who receive the pension five years earlier than do men. Even so, the self-reported health of women is worse than that of men up to age 60, akin to the situation in the US, where the two converge at age 65. Indeed, Sadana et al (2000) analyze 64 surveys of individuals from 46 countries, and find that women have worse self-reported health status in virtually all cases. That in both the US and South Africa elderly women have the advantage in health status over elderly men is consistent with (although hardly required by) their lower mortality rates.

The relationship between the South African and American data is explored numerically in Table 4. In this table, we report the results of running ordered probits on SRHS, so that our results are no longer hostage to the choice of the 1 through 5 (or any other) scale. These probits have SRHS as the dependent variable and age, sex, and race as the independent variables. Age is interacted with sex, so that we can allow for the possibility—suggested by Figures 16 and 17—that the gap between men and women's health changes with age. The first column is for the US and covers all persons, but includes an indicator for black. The second column is for the black subsample of the NHIS interviewees, still of more than 100,000 people. The last column is for the Langeberg survey, and includes a dummy variable for “coloured” compared with the omitted category, which is black.

As expected, health worsens with age. Women report themselves as being in worse health than men, and the interaction between age and female attracts a negative coefficient, which means that women age less rapidly than men, or equivalently, that the health disadvantage of women diminishes with age. In the US, blacks report worse health than do whites. If we use the effects of age to “normalize” these estimates, American blacks are nearly 18 years “older” than American whites in self-reported health terms; a 25-year old black person has the same health, on average, as a 43-year old white person. The difference between women and men in the US turns into a difference of “only” 12.5 years of (female) age for a woman aged 18, a number that diminishes to 7.6 years at age 40, and that is the same for blacks alone as for whites and blacks combined. Blacks and coloureds in South Africa age almost twice as rapidly as Americans and, as in the US, women and blacks report worse health. The effect on SRHS of being female is equivalent to 20 years in South Africa, and the effect of being black 18 years. None of these differences seem implausible, and there is no evidence in this South African to American comparison to suggest that we should be suspicious of the SRHS measures.

These results raise the further question of whether the gender difference in SRHS is conditioned by income. We know that this is sometimes the case for education, where discrimination against girls is usually more severe among the lower income groups, Filmer (1999). The South African sample size is probably too small to look at this question convincingly, but there are nearly three-quarters of a million observations in the American data and we provide some preliminary evidence in Figure 18. For each two year age band, e.g. 18 and 19 year-olds, 20 and 21 year-olds, and so on, we separate individuals into quartiles of family income within their age band. We then plot average SRHS by age and sex, with separate plots for

those in the top and bottom quartiles. As is to be expected from the health “gradient”, SRHS is worse (higher) for the “poor” men and women than for the “rich” men and women. Among the rich, average SRHS gets steadily worse with age, and women’s SRHS is always worse than men’s on average. The pattern among the poor is quite different. Women start off with a disadvantage, which gradually wears off with age, so that after about age 50, women’s SRHS is better than men’s. For both men and women in the (income-)poorer group, SRHS stops getting worse with age after age 60, and there is perhaps even somewhat of an improvement, in sharp contrast, not only with the rich men and women, but also with prior expectations. These results require more exploration than can be provided here, and similar experiments need to be carried out in other countries. However, a number of tentative hypotheses can be put forward. The much sharper deterioration in SRHS with age among the poor may come from the greater wear and tear of work in blue-collar rather than white-collar occupations, which is plausibly greater for men than for women. The change in the pattern at age 60 would then be consistent with the cessation of work, and perhaps with greater mortality selection among the poor than the rich.

Table 5 uses the Langeberg data to explore the relationship between SRHS and self-reported chronic conditions. There are eight of these, asthma, tuberculosis, cancer, heart trouble, stroke, high cholesterol, diabetes, and emphysema, each of which was reported in answer to questions of the form, “Have you ever been told by a doctor, a nurse, or a healthcare professional that you have X?” The table reports the results from a single ordered probit containing, in addition to age, a dummy for female, a dummy for race, an age-female interaction, and dummies for the presence of each of the conditions, with each effect allowed to be different for men and women. Our aim here is twofold, first to see whether the inferior SRHS of women can be attributed to their having

more health conditions and, second, to see whether men and women's SRHS responds differently to these medical conditions.

The answer to the first question is no. Even though the presence or absence of conditions has a large and significant effect on SRHS (the  $\chi^2$  tests for their absence are 119.6 for males and 606.2 for females), adding them to the probit, even with different coefficients for men and women, makes the estimated female dummy only slightly smaller than it was in Table 4. Of the conditions themselves, asthma, tuberculosis, heart trouble, high cholesterol, and diabetes all sharply diminish the self-reported health-status of the people who report having been told they have these conditions. Cancer has only an insignificant effect, as does emphysema; both are conditions that are much more common among the white population than among the black and coloured populations analyzed here. For the five conditions that have significant effects, the effects on men's self-reported health status is always similar to the effects on women's self-reported health status, and in none of these cases can we reject the hypothesis that the coefficients are the same by sex. Only for the cancer, where the effects are insignificant for men and women separately, can we marginally reject the hypothesis that the effects are the same.

Medically diagnosed conditions cause people to report lower self-reported health and the effects on health status of common, important conditions are the same for men and women. The causes of women reporting worse health than men lie elsewhere than in the medical conditions with which they have been diagnosed.

Table 6 repeats the analysis of Table 5 but using ADLs and IADLs rather than professionally diagnosed conditions. Because the Langeberg asked questions about activity limitations only for people aged 55 and over, there are only 67 people in the sample used for this table. Nevertheless,

the limitations are associated with worse self-reported health status, especially difficulties with dressing (for men), bathing, walking (for men), lifting heavy objects, and light housework (for women.) Contrary to what we find for the professionally diagnosed conditions, limitations in activities of daily living have quite different effects on the self-reported health status of men and women. For example, difficulty with light housework makes women's SRHS worse but, if anything, *improves* the SRHS of men, opening an interesting window on gender roles in South Africa. Difficulties with walking has negative health consequences for men, but not for women. Obesity is common among of these South African women, but is frequently not seen as a problem; indeed, the inability to get around and to be ministered to by others is often seen as a mark of distinction and respect. Having to be helped with dressing may be a limitation that is similarly differentiated by gender, especially in large households where there are other people to help.

We find the results in Table 6 plausible on their own terms. The way in which ADLs and IADLs affect health status is structured according to gender roles in daily living. Similar results were found in Indonesia by Frankenberg and Thomas (2000), who also noted an effect of economic status; high status women, who do not regard certain activities as part of their lives, react very differently to limitations on those activities than do the poorer women who routinely perform them. Such findings are supportive of the usefulness of ADL and IADL measures for their own purposes, and are consistent with our working hypothesis that self-reported health status contains useful information about self-perceived health. Of course the findings militate *against* the use of functional limitations to measure well-being and poverty, however important they may be in helping to explain labor market participation or other outcomes.

### **2.3 Problems with self-reports: the WHO studies**

In a series of recent studies, directed by Chris Murray, the World Health Organization has embarked on a large-scale household survey program to measure non-fatal health. Although concerned only with the health dimension of poverty, this program has much in common with the more general measurement of wellbeing and poverty that is the topic of this paper. In this final section, we draw some lessons from their work, and make some recommendations for the kind of work that the World Bank might undertake in order to collect better data on poverty and gender.

Sadana et al (2000) report the analysis of self-reported health information from 64 pre-existing surveys from 46 countries. These surveys take a wide-range of approaches, including the collection of self-reported health status as used above, but some also gather reports of illness, of injury, as well as asking people about ADLs and IADLs. The overarching aim of the WHO project is to measure non-fatal health from household surveys in a manner that yields comparable results across countries. Although international comparability is neither necessary nor sufficient for comparability by gender many of the same concerns arise for both. Sadana et al. make the assumption that the various reports in the various surveys are all attributable to some latent, underlying health status variable, and attempt to recover it in a comparable way for all the surveys and countries. They assess this exercise as a failure, and for the same reasons that earlier measures, typically illness measures, have not worked: the responses to the same objective state of health are conditioned by the different characteristics of the respondents, leading to implausible cross-country patterns of disease or, within countries, to clearly incorrect relationships between socioeconomic status and health. Faced with this failure, the project has designed a new methodology, the much superior results of which are reported in Sadana et al

(2002). On this, more below.

Note first that not all of the problems in the analysis concern the SRHS measure used in Section 2.2. It has been known for a long time that self-reports of sickness and injury are inappropriately conditioned by individual circumstance, and it is no surprise that the health consequences of ADLs and IADLs are different under different circumstances. Indeed, there are strong theoretical grounds for *not* treating health as separable from consumption and production activities, see Broome (2001). That said, Sadana et al. do identify problems with the SRHS measures, which show differences in distributions of self-reported health across countries that cannot be readily attributed to real differences in health. We note, however, that their results show consistently worse health for women across the countries in their survey. In consequence, even if it was necessary to standardize SRHS for each country, an expedient that would preclude international comparisons, the measure would still be useful for exploring variations in health by gender within each country.

The WHO team proposes a theoretical framework for thinking about non-comparability of health reports. To simplify, we can think of SRHS as the result of an ordered probit applied to an underlying level of health status, while recognizing that the cut-points of the probit, the values at which people switch from one class to the next, from “fair” to “poor,” for example, also vary from individual to individual. Just as in Tables 4 and 5, we can model individual health as a function of individual characteristics, such as sex and age, but we must also allow the cut-points to be a function of the same, or other, characteristics. A poor person may have a higher threshold for pain, and so will only report his health as “poor” in much more dire circumstances than for a rich person. Of course, such a model is unidentified. We have no way of telling whether some

individuals report that they are in pain because they are indeed in pain, or because they have a low pain threshold. To solve this problem, the WHO team introduce a series of questions that allow the respondent to identify the way he or she perceives health, independently of his or her own health status. This is done using a series of “vignettes,” in each of which the respondent is presented with the description of a fictional person, whose activities, symptoms, abilities, and general appearance are described, and whose health is rated by the respondent on the standard Likert scale. These vignettes effectively establish the scale which the respondent uses, and allows the investigators to “self-standardize” each individual response.

Sadana et al (2002) presents the results of using this procedure in 66 surveys in 57 countries. Many of the evident problems of international comparability in the earlier study are now resolved or ameliorated, and certainly the new health measures appear to behave in a much more reasonable way than the old. Even so, they retain the property of the unadjusted measures that women’s health is consistently poorer than men’s health. Because this work is so recent, there may possibly be problems that have not yet been identified. In particular, there is relatively little experience with these kinds of vignettes, and the extent to which they solve the problems of position affecting perception. For example, it is possible that people’s own health might affect their perception of the health of others, the “you cannot really sympathize unless you are sick yourself” syndrome. Nevertheless, and even if, as we believe, current SRHS measures contain valuable information on health poverty by gender, the use of vignettes in new surveys is likely to improve comparability by gender, or to least to offer some insurance against skeptics.

Even so, the new WHO surveys are not suitable for poverty analysis as they stand. Just as economic household surveys fail by not gathering health information, so do the WHO surveys

fall short by not collecting economic data. Separate surveys cannot serve the purpose because the different measures of wellbeing are correlated; poor people tend to do worse on all measures, and possibly poor women worse of all. So if the World Bank is serious about measuring poverty in a broader way, it needs a new (but not necessarily very complicated nor long) survey that collects, (a) self-reported health measures, together with the WHO vignettes, (b) a minimal list of consumption items, including twenty to thirty items, and in some cases, a small number of income questions, and (c) standard questions on education and literacy. Armed with such an instrument, we would be much better able to measure poverty, including its gender dimension.

### **3. Conclusions and recommendations**

Standard poverty measures, used by the World Bank and others, are *income* measures based on the adequacy of household income or consumption. If we are to disaggregate such measures by gender, we need some imputation method for allocating consumption, income, or wellbeing, to individual household members. While there exist a number of such methods in the literature, none has commanded the broad acceptance that would be required before it could be routinely used to measure income poverty by gender. Such a conclusion does not mean that the topic should be abandoned, just that we are still at the stage of research, not production. As far as research is concerned, there is a good case for collecting more consumption data on an individual basis. There are some private goods, such as school fees, medical expenses, or tobacco consumption, whose assignment to individuals is relatively straightforward, and where the information is of interest in its own right. Moreover, there are many countries where the consumption of some items by household members is private information that is not shared with

other members of the household, so that interviewing individuals about their own consumption, at least for some items, would improve the accuracy of the household consumption totals that are currently of primary interest for welfare and poverty measurement. Such information will also provide a platform that will encourage researchers to develop methods for examining intrahousehold allocation.

It is also possible that income is not the right space in which to look for gender differences in poverty. Even when gender differences in consumption are identified, it is always going to be difficult to tell whether these correspond to real differences in well-being. For example, we can imagine a “traditional” family structure in which men get more than their share of food and of spending money for leisure activities, but in which they also do more than their share of the work, but where women have almost all the power in domestic arrangements. Such an arrangement might be discriminatory and unfair, but it also might not be. There is therefore everything to be said for moving away from income to measures that are broader, including health and education, and that have the further advantage of being immediately measurable on an individual basis.

To this end, we believe that the those interested in surveys of wellbeing should aggressively pursue the collection of self-reported health information. This is currently an active area of research and, although there are skeptics, our reading of the literature is that the differences in reports by men and women are sufficiently consistent across studies to suggest that they are picking up a real difference in perceived health. The WHO studies contain a great deal of information on this that is not immediately accessible, because the main focus of the reports is on international comparisons, not gender comparisons. Nevertheless, the information is presumably

available for further research. Note also that the standard self-reported health measures are quick and easy to answer, so that there is a strong case for their routine inclusion in consumption and income surveys such as the LSMS surveys. This is very much in line with other current work exploring other ordinal measures of wellbeing, particularly on poverty and on happiness.

The collection of broader measures of wellbeing does not mean the abandonment of traditional income measures. Indeed, the reverse is true. The different aspects of well-being, income, educational, and health, are not distributed independently of one another. Those who are income-deprived are typically more likely to be deprived in terms of health and education. In consequence, it is of the greatest importance that we collect data on all the dimensions of well-being simultaneously. Otherwise we understate the depth of poverty and understate inequalities between rich and poor. So our survey instruments must not focus on income, health, or education, but collect data on all from the same people.

## List of works cited:

- Belcher, D. W., A. K. Neumann, F. K. Wurapa, and I. M. Lourie, 1976, "Comparison of morbidity interviews with a health examination survey in rural Africa," *American Journal of Tropical Medicine and Hygiene*, **25**(5), 751–8.
- Bidani, Benu and Martin Ravallion, 1997. "Decomposing social indicators using distributional data," *Journal of Econometrics*, **77**, 125-139.
- Bourguignon, François and Pierre-André Chiappori, 1992, "Collective models of household behavior: an introduction," *European Economic Review*, **36**, 355–64.
- Broome, John, 2002, "Measuring the burden of disease by aggregating well-being," Chapter 3.1 in Christopher J. L. Murray, J. A. Saloman, C. D. Mathers, and A.D. Lopez, eds., *Summary measures of population health*, Geneva. World Health Organization.
- Browning, Martin, François Bourguignon, Pierre-André Chiappori, and Valérie Lechene, 1994, "Income and outcomes: a structural model of intrahousehold allocation," *Journal of Political Economy*, **102**, 1067–96.
- Case, Anne C., 2002, "Health, income, and economic development," *Annual World Bank Conference on Development Economics 2001/2002*, Washington, DC. The World Bank, 221–41.
- Chiappori, Pierre-André, 1988, "Rational household labor supply," *Econometrica*, **56**, 63–89.
- Chiappori, Pierre-André, 1992, "Collective labor supply and welfare," *Journal of Political Economy*, **100**, 437–67.
- Chesher, Andrew. 1997. "Diet revealed?: semiparametric estimation of nutrient intake-age relationships," *Journal of the Royal Statistical Society A*, **160**(3):389-428.
- Clark, Shelley, 2000, "Son preference and sex composition of children: evidence from India," *Demography*, **37**, 95–108.
- Das Gupta, Monica, 1987, "Selective discrimination against female children in rural Punjab, India," *Population and Development Review*, **13**, 77–100.
- Deaton, Angus, 1989, "Looking for boy-girl discrimination in household expenditure data," *World Bank Economic Review*, **3**, 1–15.
- Deaton, Angus, 1997, *The analysis of household surveys: a microeconomic approach to development policy*, Johns Hopkins University Press.

Deaton, Angus and Christina Paxson, 1998, "Economies of scale, household size, and the demand for food," *Journal of Political Economy*, **106**, 897–930.

Deaton, Angus and Christina Paxson, 2000, "Growth and saving among individuals and households," *Review of Economics and Statistics*, **82**(2), 212–25.

Deaton, Angus and Christina Paxson, 2001, "Poverty among the elderly and children in developing countries," Research Program in Development Studies, Princeton University, processed.

Deaton, Angus, Javier Ruix-Castillo, and Duncan Thomas, 1989, "The influence of household composition on household expenditure patterns: theory and Spanish evidence," *Journal of Political Economy*, **97**, 197–200.

Filmer, Deon, 1999, "The structure of social disparities in education: gender and wealth," Policy Research Report on Gender and Development, Working Paper No. 5, Washington, DC, The World Bank.

Idler, E. L., and Y. Benjamini, 1997, "Self-rated health and mortality: a review of twenty-seven community studies," *Journal of Health and Social Behavior*, **38**(1), 21–37.

Mason, Andrew D., Claudia E. Montenegro, and Shahidur R. Khandker, 1999, "Can we say anything more about gender and poverty using household consumption data," Poverty Reduction and Economic Management Network, The World Bank, processed. (July.)

Murray, Christopher J. L., and Lincoln Chen, 1992, "Understanding morbidity change," *Population and Development Review*, **18**(3), 481–503.

Murthi, Mamta, Anne-Catherine Guio, and Jean Drèze, 1995, "Mortality, fertility, and gender bias in India," *Population and Development Review*, **21** (4), December.

Pollak, Robert A. and Terence J Wales, 1979, "Welfare comparisons and equivalence scales," *American Economic Review*, **69**(2), May, 216–21.

Rothbarth, Erwin, 1943, "Notes on a method of determining equivalent income for families of different composition," Appendix 4 in C. Madge, ed., *War time pattern of saving and spending* Occasional Paper No. 4, London, National Institute of Economic and Social Research.

Sadana, Ritu, Colin D. Mather, Alan D. Lopez, Cheistopher J. L. Murray, and Kim Iburg, 2000, "Comparative analysis of more than 50 household surveys on health status," GPE Discussion Paper Series, No. 15, Geneva. World Health Organization.

Sadana, Ritu, Ajay Tandon, Christopher J. L. Murray, Irina Serdobova, Yang Cao, Wan Jun Xie,

Somnath Chatterji, and Bedirhan L. Ustün, 2002, “Describing population health in six domains: comparable results from 66 households surveys,” GPE Discussion Paper Series, No. 43, Geneva. World Health Organization.

Sen, Amartya K., 2002, “Health: perception versus observation,” *British Medical Journal*, **324**, 860–1.

Thomas, Duncan and Elizabeth Frankenberg, 2000, “The measurement and interpretation of health in social surveys,” Labor and Population Program, Working Paper 1–06, Santa Monica, Ca. RAND Corporation

Van Praag, Bernard M S., and Marcel F Warnaar, 1997, “The cost of children and the use of demographic variables in consumer demand,” in Mark R. Rosenzweig and Oded Stark, editors, *Handbook of Population and Family Economics*, Volume 1A. Amsterdam. Elsevier.

Yamaguchi, K., 1989, “A formal theory for male-preferring stopping rules of child-bearing,” *Demography*, **26**(3), 451–65.

**Table 1**  
**Per capita total household expenditures and headcount poverty rates, rural India 1999–2000**

	Average per capita total expenditure			Head count ratios		
	All	Males	Females	All	Males	Females
Andhra Pradesh	453.48	454.53	452.43	10.53	10.43	10.63
Assam	425.21	428.18	421.84	40.27	39.34	41.32
Bihar	384.68	386.67	382.52	43.98	43.31	44.70
Gujarat	551.24	555.65	546.36	12.36	11.99	12.76
Haryana	714.16	714.01	714.33	7.41	7.41	7.39
Himachal Pradesh	688.83	697.71	680.16	7.47	7.01	7.92
Karnataka	499.98	504.39	495.27	16.84	16.60	17.11
Kerala	766.15	776.33	756.68	9.37	8.73	9.97
Madhya Pradesh	401.35	407.76	394.60	37.25	36.36	38.19
Maharashtra	496.60	503.70	489.35	23.22	22.59	23.85
Orissa	373.94	380.50	367.37	47.79	46.58	49.00
Punjab	742.30	753.41	730.06	5.99	5.62	6.40
Rajasthan	548.78	553.36	543.86	13.47	12.92	14.07
Tamil Nadu	513.75	522.74	504.67	20.02	19.21	20.84
Uttar Pradesh	466.43	470.14	462.38	31.06	30.41	31.77
West Bengal	454.27	455.31	453.16	31.66	30.85	32.53
All India	485.87	490.39	480.05	26.77	26.22	27.35

*Notes:* Authors' calculations from the NSS 55<sup>th</sup> Round, rural India, 1999–2000. The numbers are calculated as follows. Each person in each household is assigned the pce level of the household, as well as an indicator of whether or not the household is in poverty. These individual pce levels and poverty indicators are then averaged over males and females separately, and over all persons. Household inflation factors supplied by the NSS are used. The Planning Commission's poverty lines for the 55<sup>th</sup> Round are used, so that the headcount ratios in the "All" column correspond (up to rounding error) to the official poverty counts.

**Table 2**  
**Average household size by gender and age**  
Rural India, 1999–2000

	Males	Females	Males 0–14	Females 0–14
Andhra Pradesh	5.07	5.02	5.36	5.52
Assam	6.48	6.50	6.65	6.74
Bihar	6.82	6.76	7.17	7.22
Gujarat	6.09	6.21	6.45	6.78
Haryana	6.49	6.62	6.55	6.88
Himachal Pradesh	5.75	5.88	6.15	6.40
Karnataka	6.18	6.13	6.55	6.66
Kerala	5.43	5.40	5.80	5.79
Madhya Pradesh	6.70	6.82	6.97	7.28
Maharashtra	5.81	5.91	6.09	6.41
Orissa	5.75	5.79	6.07	6.33
Punjab	6.51	6.68	6.75	7.00
Rajasthan	7.01	7.10	7.25	7.46
Tamil Nadu	4.85	4.79	5.25	5.29
Uttar Pradesh	7.25	7.33	7.60	7.72
West Bengal	6.19	6.17	6.36	6.59
All India	6.33	6.35	6.72	6.88

*Notes:* The table shows the average number of people who share a household with the reference person, including him or herself, distinguished by sex and age in the columns. The numbers are computed by assigning to each person in the survey his or her household size, and then averaging over all males, females, boys, and girls. Note that these numbers are larger than average household size, because they are computed on an individual, not a household basis; the individual basis numbers can be obtained as household averages if the latter are weighted by household size.

**Table 3**  
**Adult goods regressions: rural India, 1999–2000**

	ALL ADULT GOODS	TOBACCO	PAN	INTOXI- CANTS
ln per capita expenditure	0.330 (8.9)	-0.083 (3.9)	-0.042 (3.0)	0.456 (18.2)
ln household size	-0.030 (0.9)	-0.120 (6.4)	0.021 (1.7)	0.069 (3.1)
ratio of males aged 0–2	-0.227 (1.0)	0.767 (5.7)	-1.278 (14.5)	0.284 (1.8)
ratio of males aged 3–4	0.035 (0.1)	0.791 (5.5)	-1.256 (13.2)	0.501 (3.0)
ratio of males aged 5–14	-0.377 (2.8)	0.441 (5.7)	-1.191 (23.6)	0.374 (4.2)
ratio of males aged 15–55	1.556 (13.0)	1.650 (23.9)	-0.853 (18.8)	0.760 (9.4)
ratio of males 56 plus	2.276 (13.8)	2.352 (24.9)	-0.864 (13.8)	0.788 (7.1)
ratio of females aged 0–2	-0.109 (0.5)	0.923 (6.8)	-1.213 (13.5)	0.182 (1.1)
ratio of females aged 3–4	0.469 (1.8)	1.101 (7.4)	-1.271 (12.9)	0.639 (3.7)
ratio of females aged 5–14	-0.195 (1.3)	0.547 (6.6)	-1.178 (21.4)	0.436 (4.5)
ratio of females aged 15–55	0.612 (5.2)	-0.208 (3.1)	-0.727 (16.4)	0.323 (4.1)

*Notes:* Controls are also included (but not shown) for household type (self-employed outside of agriculture, agricultural laborer, non-agricultural laborer, self-employed in agriculture, and other), caste (schedule tribe, scheduled caste, other backward caste, other), and religion (hindu versus other). The sex and age variables are entered as ratios relative to total household size, so that, for example, females 5–14 is the ratio of the number of females aged 5 to 14 to total household size. Females aged 55 and above are the omitted category. The budget shares are the dependent variables and are expressed as percentages. *pce* is total household expenditure per capita, and *hhsiz*e is household size; households with zero budget shares are included in the regressions. The figures in parentheses are (absolute) *t*-values. There are 71,182 observations in each regression.

**Table 4. Health Status of Women and Men by Age, South Africa and US Data**

	US 1986-95 All	US 1986-95 Blacks only	Langeberg 1999 Blacks and Coloureds
Female	.297 (.007)	.323 (.018)	.818 (.284)
Female×Age	-.004 (.0001)	-.004 (.0004)	-.016 (.006)
Age	.022 (.0001)	.024 (.0003)	.040 (.005)
Indicator: Black	.395 (.004)	--	--
Indicator: Coloured	--	--	-.721 (.152)
Number of Observations	796294	107085	519

*Notes:* Ordered probit estimates, with standard errors in parentheses. Estimation of standard errors in the Langeberg survey allows for correlation in unobservables for individuals drawn from the same cluster. Data for the U.S. is taken from annual National Health Interview Surveys 1986-1995.

**Table 5. Health Status and Chronic Conditions, Black and Coloured Adults**

		Probit coefficient on Males:	Probit coefficient on Females	Chi-square test: coefficients for males =females
Female	.714 (.326)			
Female×Age	-.015 (.008)			
Age	.030 (.005)			
Condition:				
Asthma		.820 (.317)	.772 (.208)	0.02 (.9004)
TB		.709 (.411)	.842 (.300)	0.23 (.6298)
Cancer		-.525 (.398)	.373 (.297)	3.90 (.0482)
Heart Trouble		.579 (.184)	.491 (.226)	0.07 (.7986)
Stroke		-.830 (.532)	-.069 (.254)	1.68 (.1951)
High Cholesterol		.618 (.141)	.574 (.196)	0.06 (.8116)
Diabetes		.254 (.116)	.500 (.280)	0.80 (.3722)
Emphysema		.368 (.264)	.170 (.305)	0.23 (.6280)
		Chi-square test of joint significance of conditions for males:	Chi-square test of joint significance of conditions for females:	
Joint significance of conditions by sex (p-value)		119.55 (.0000)	606.17 (.0000)	

*Notes:* Ordered probit estimates, with standard errors in parentheses. Estimation of standard errors allows for correlation in unobservables for individuals drawn from the same cluster. Number of observations = 501. Also included is an indicator for race.

**Table 6. Health Status and Activities of Daily Living, Black and Coloured Elderly Adults**

		Probit coefficient on Males:	Probit coefficient on Females	Chi-square test: coefficients for males =females
Female	7.64 (4.60)			
Female×Age	-.144 (.067)			
Age	.086 (.049)			
Individual reports difficulty:				
Dressing		2.72 (.733)	-2.09 (.785)	14.80 (.0001)
Bathing		.911 (.262)	1.46 (.373)	1.79 (.1804)
Taking a bus alone		.028 (.579)	-1.42 (.761)	1.89 (.1690)
Walking		1.33 (.545)	-.103 (.390)	4.59 (.0321)
Climbing stairs		-2.30 (.818)	.541 (.476)	6.36 (.0117)
Lifting heavy objects		.865 (.733)	1.60 (.526)	0.96 (.3282)
Light housework		-1.52 (.602)	2.01 (.607)	13.83 (.0002)
Managing money		.242 (.368)	-.012 (.674)	0.13 (.7236)
		Chi-square test of joint significance of conditions for males:	Chi-square test of joint significance of conditions for females:	
Joint significance of conditions by sex (p-value)		58.57 (.0000)	123.77 (.0000)	

*Notes:* Ordered probit estimates, with standard errors in parentheses. Estimation of standard errors allows for correlation in unobservables for individuals drawn from the same cluster. Number of observations = 67. Also included is an indicator for race.

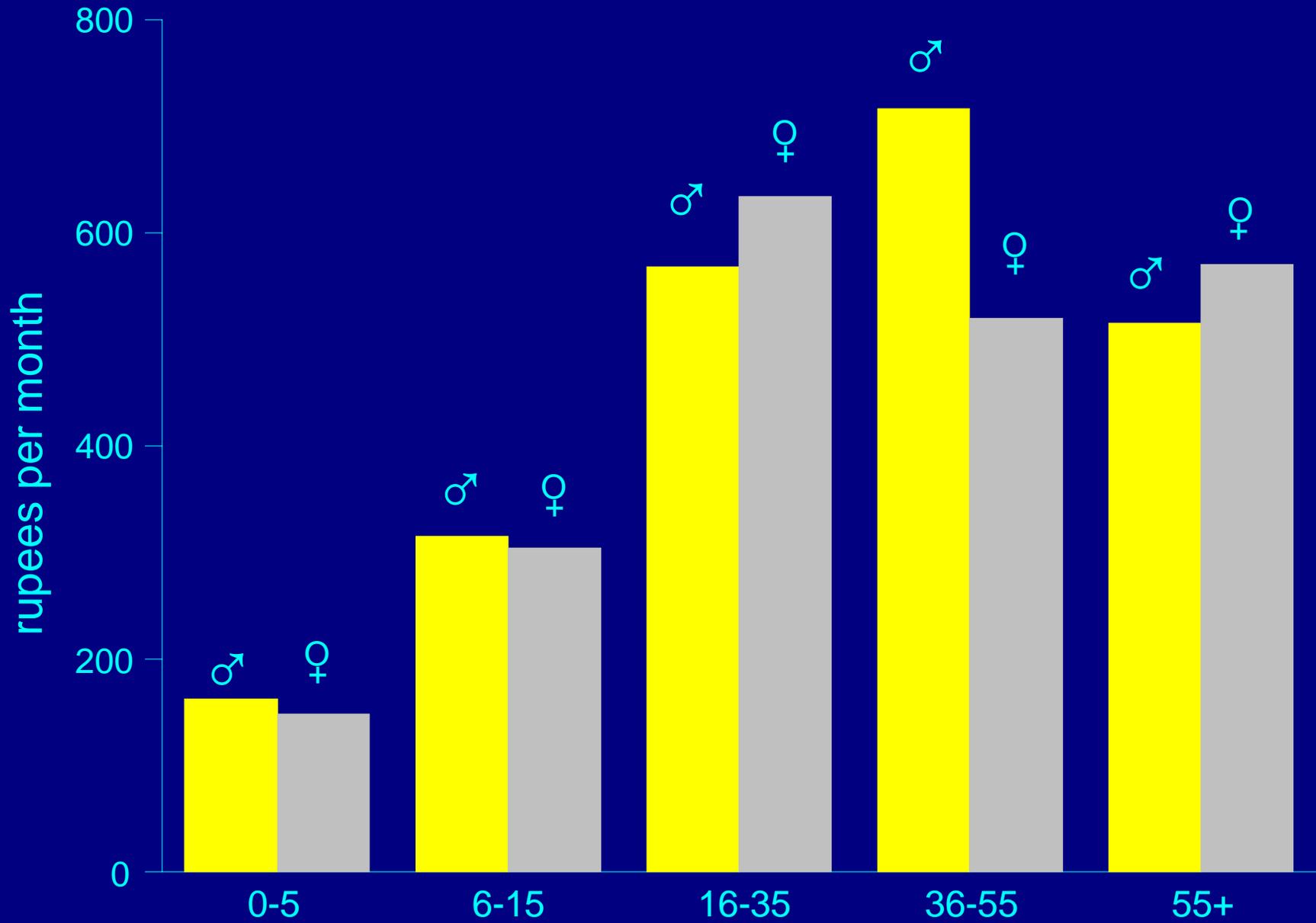


Figure 1: Total consumption, age, and gender, rural India

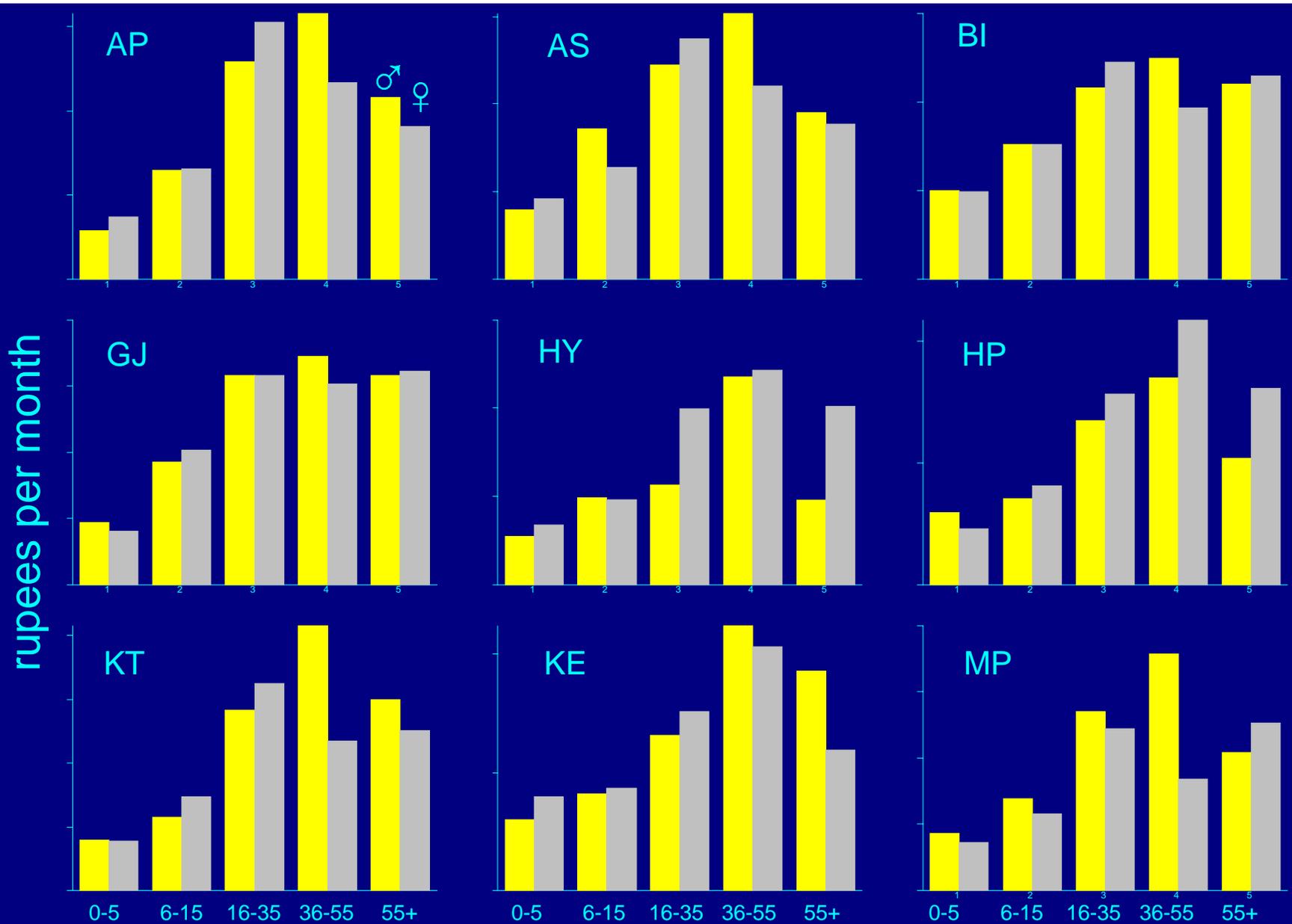


Figure 2: Total consumption, age, and gender, rural India

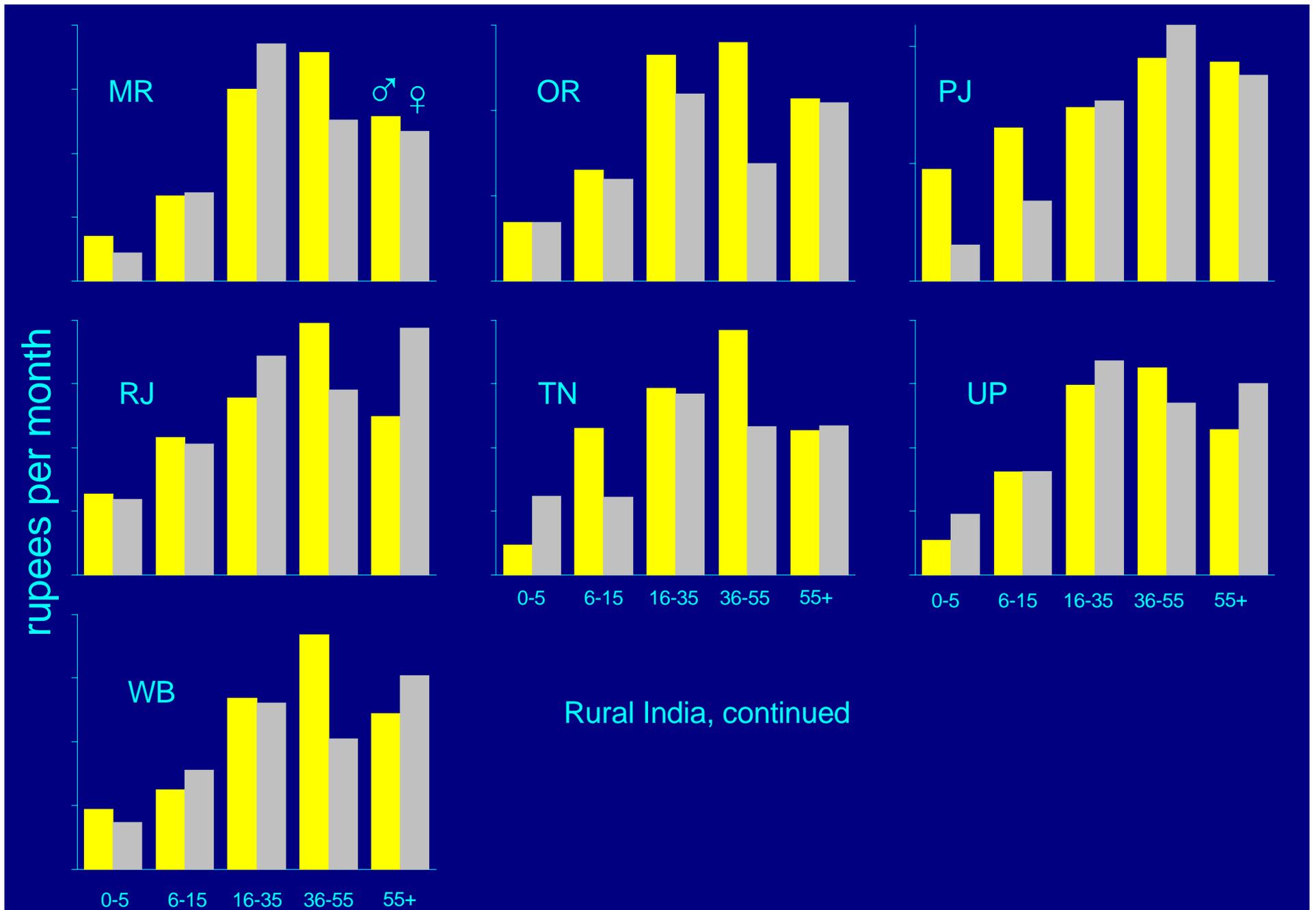
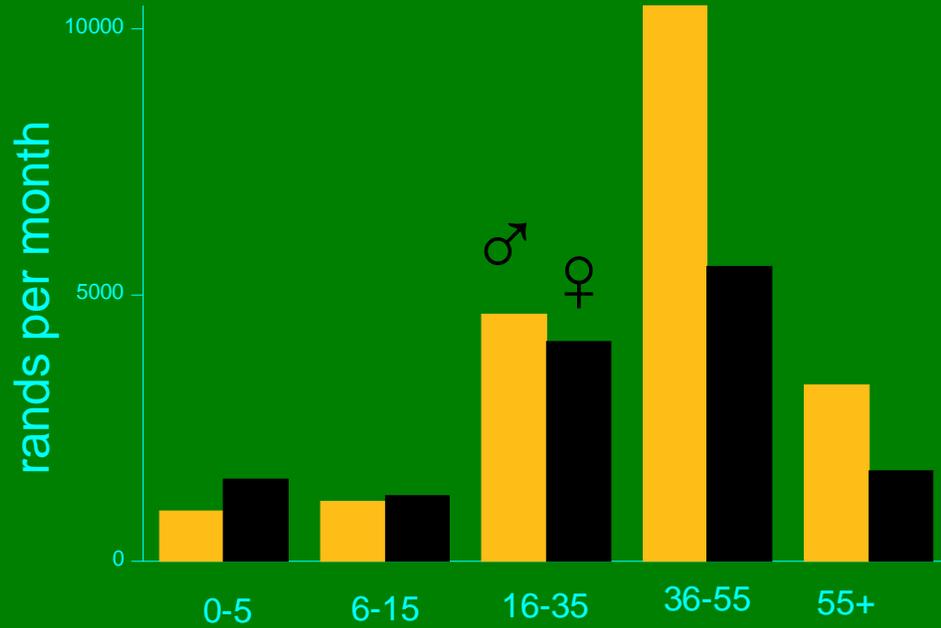
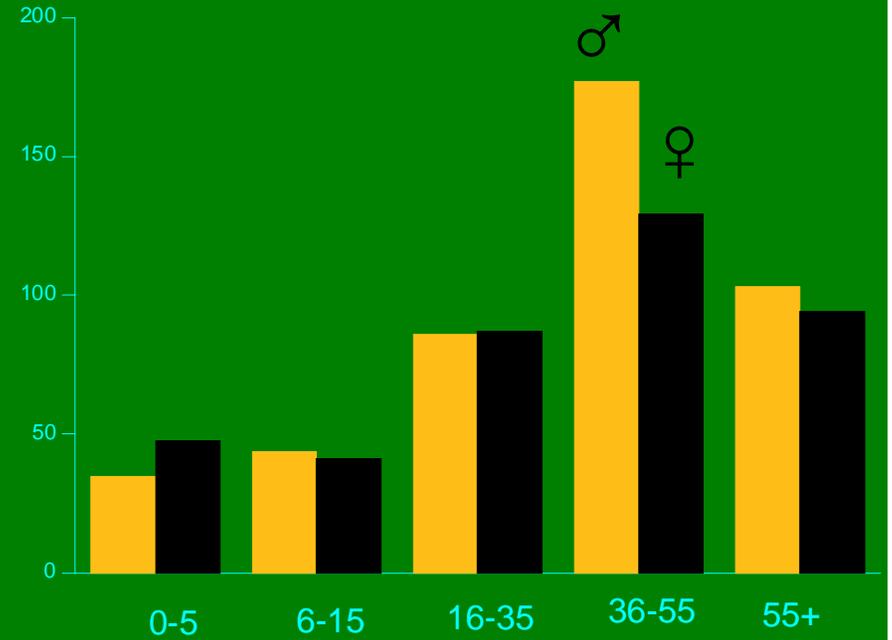


Figure 3: Total consumption, age, and gender, rural India



total expenditures



food expenditures

South Africa

Figure 4: Total household and total food expenditure

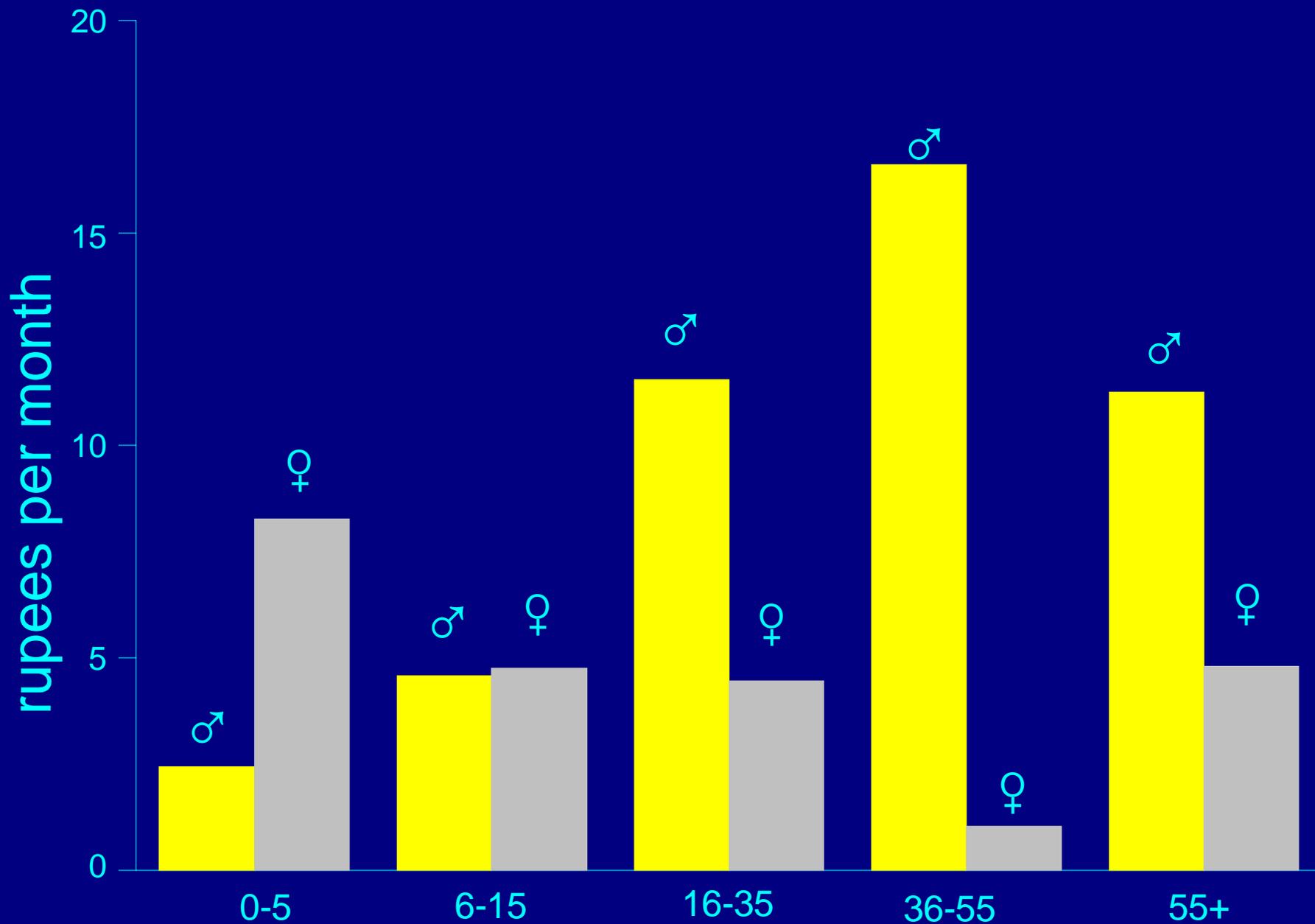


Figure 5: Tobacco consumption, age, and gender, rural India

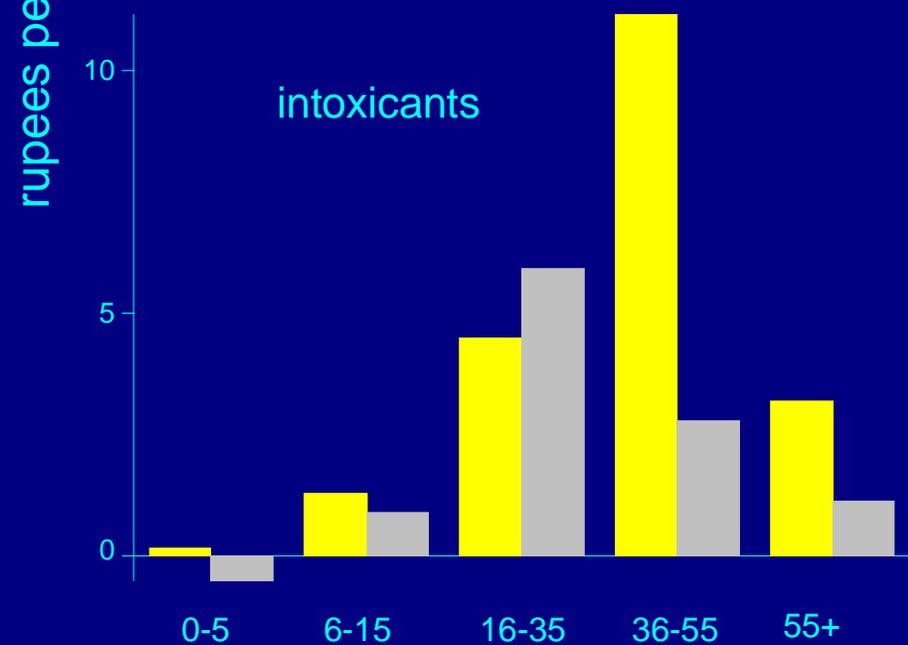
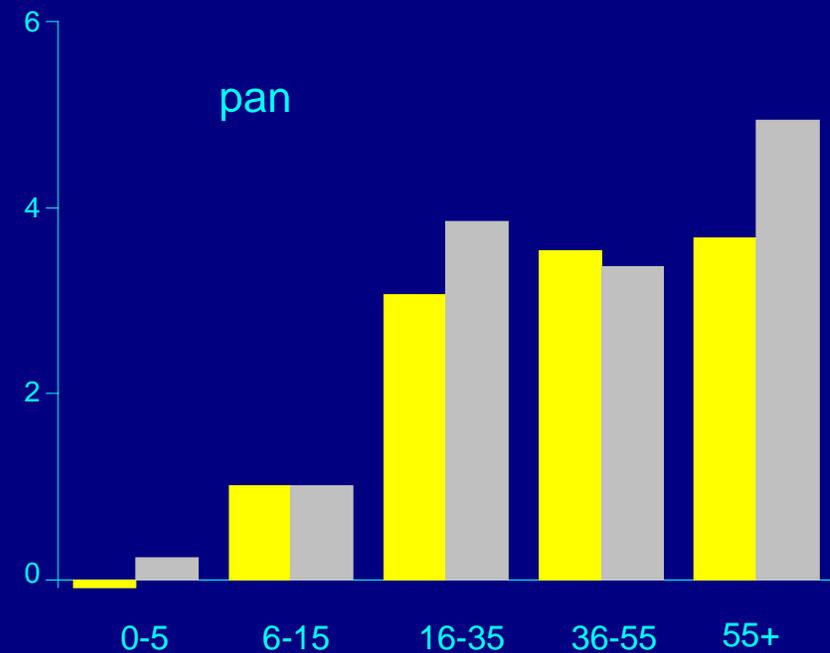
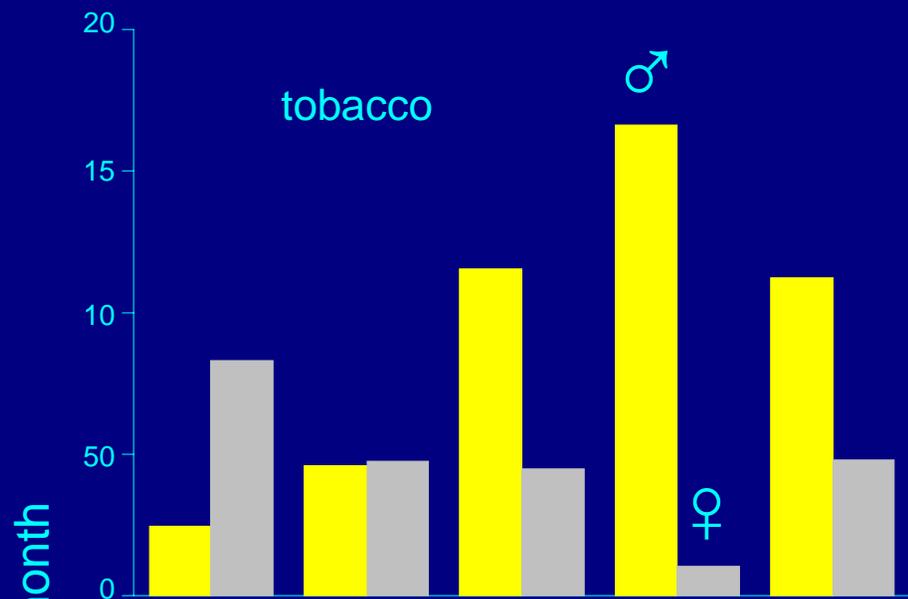


Figure 6: Adult goods, rural India

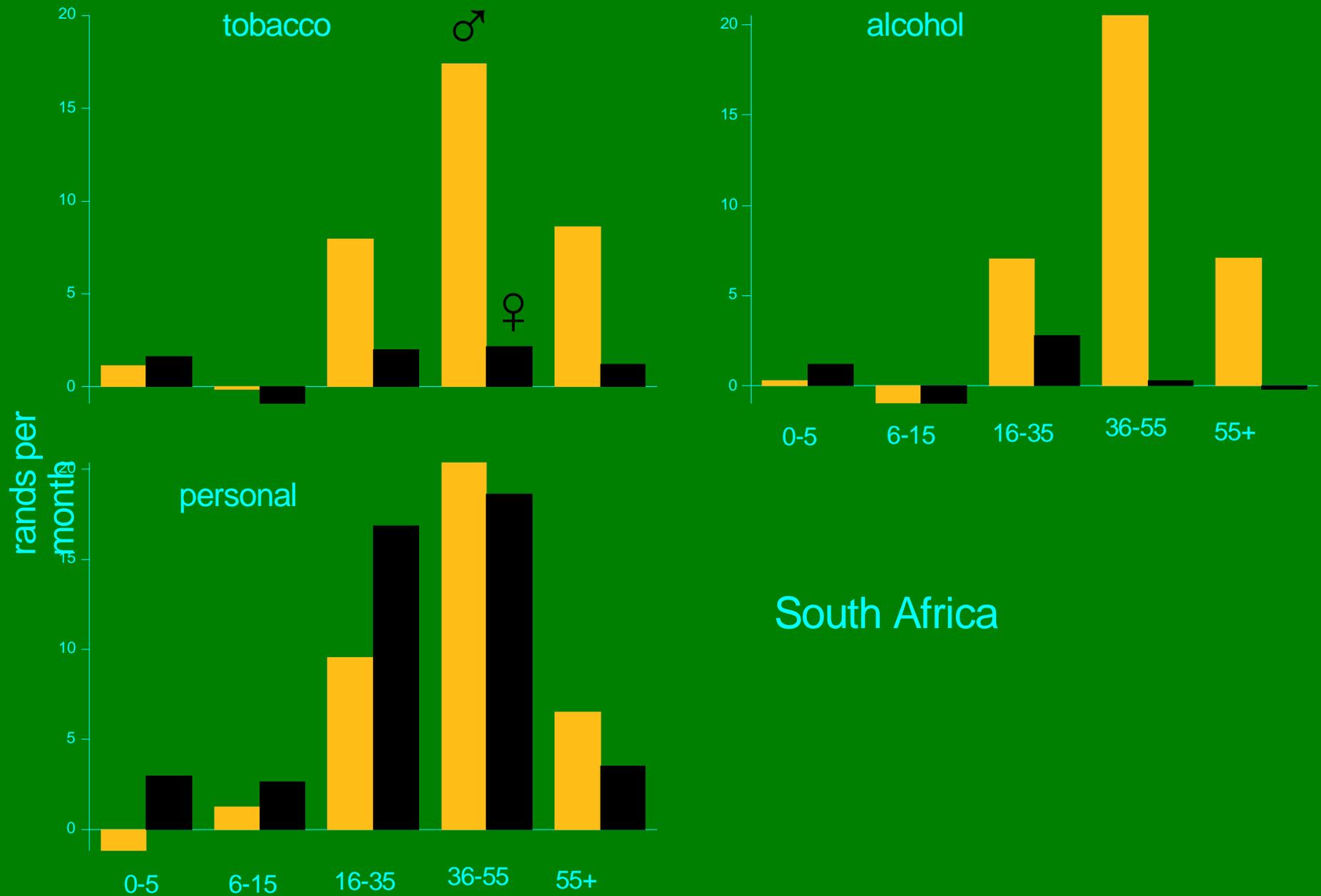


Figure 7: Tobacco, alcohol and personal expenditures

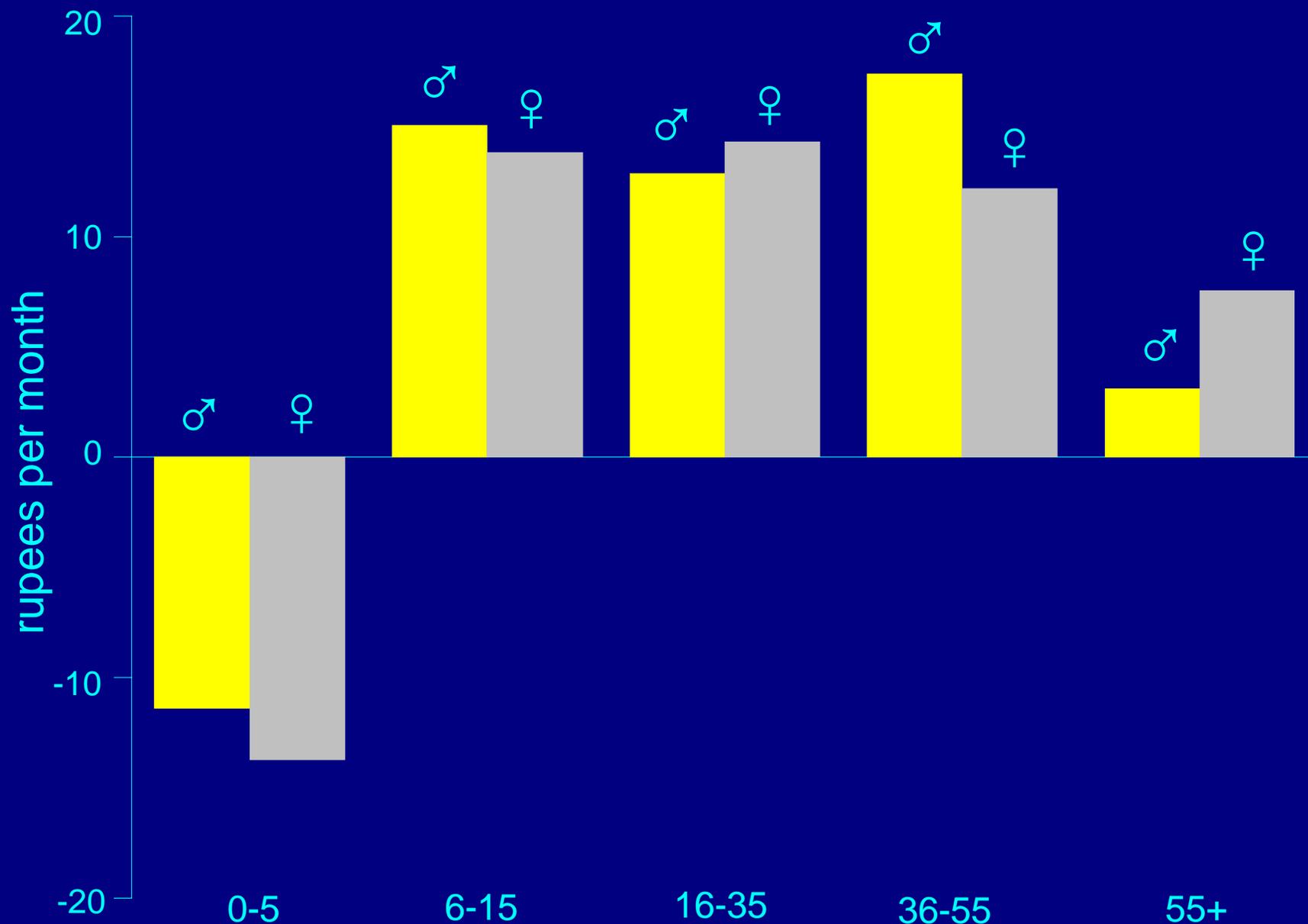
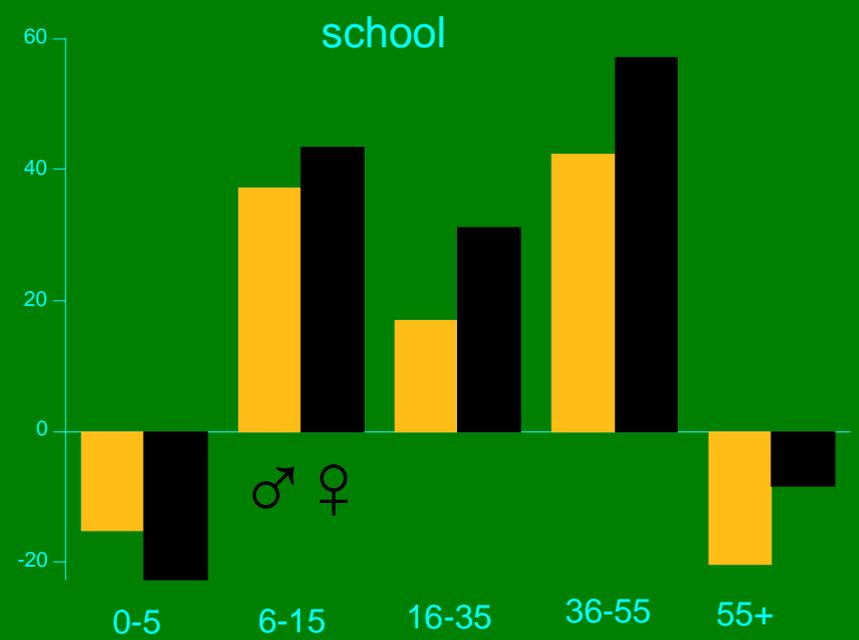
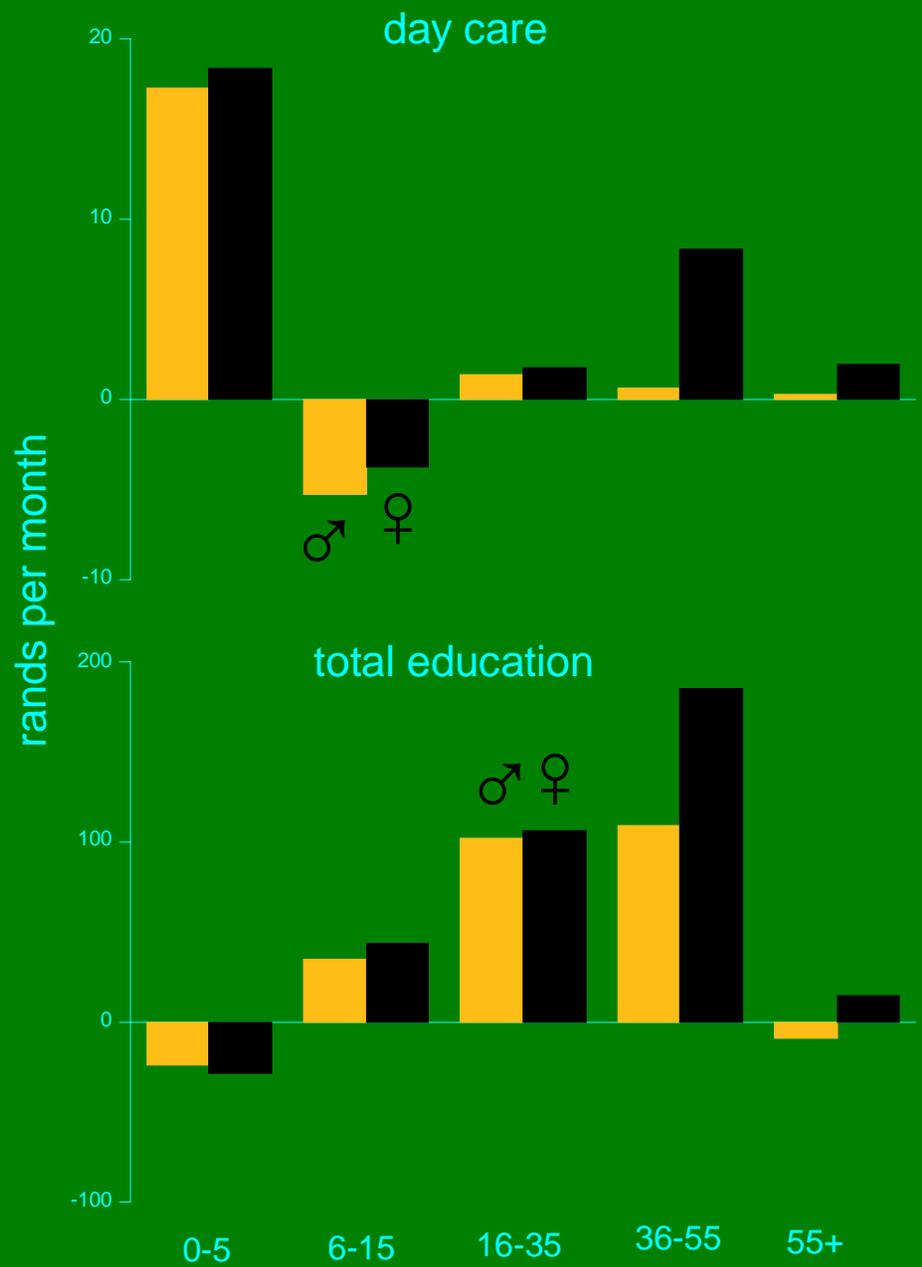


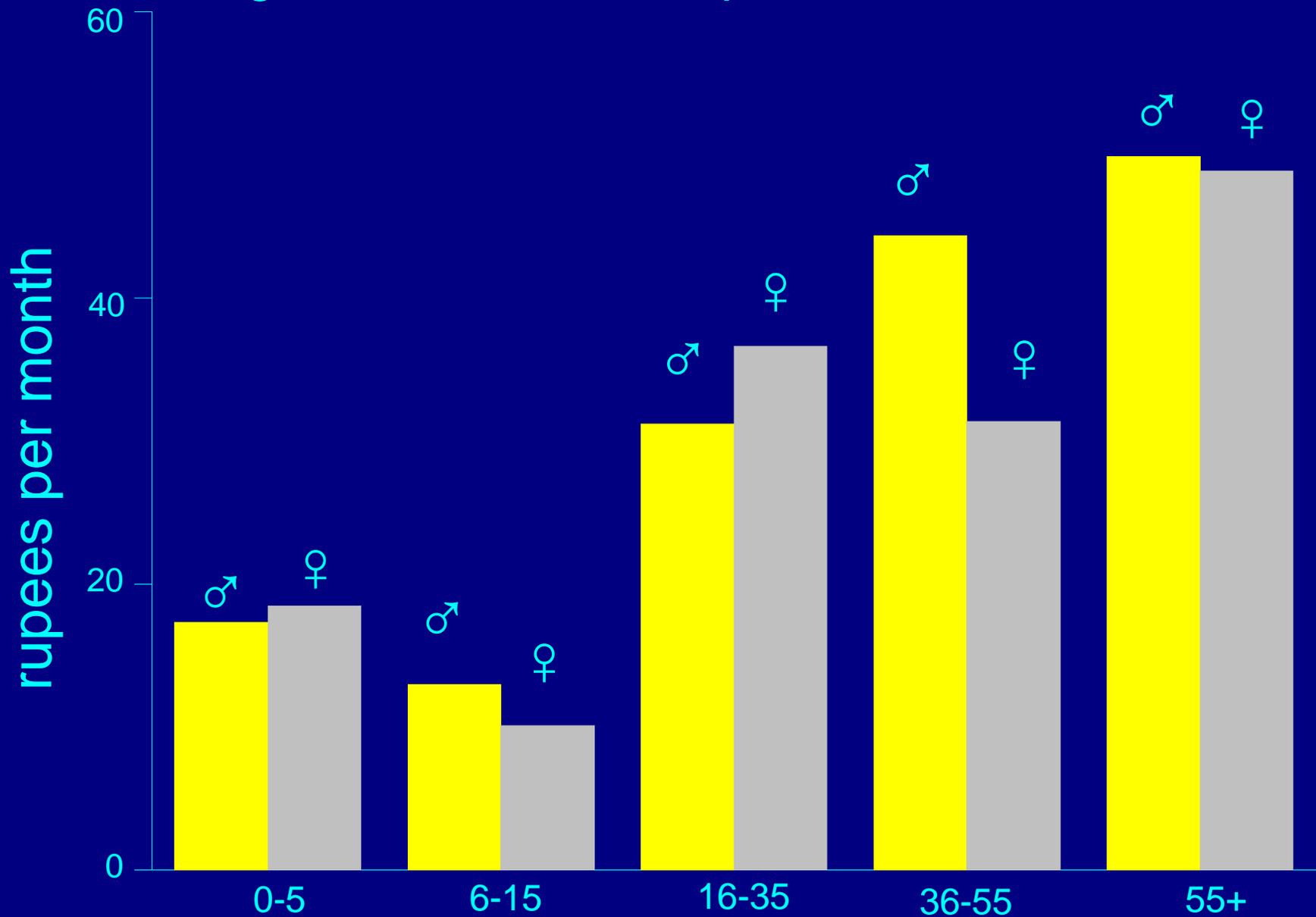
Figure 8: Educational expenditures, rural India



South Africa

Figure 9: Day care, school and total education expenditures

Figure 10: Medical expenditures, rural India



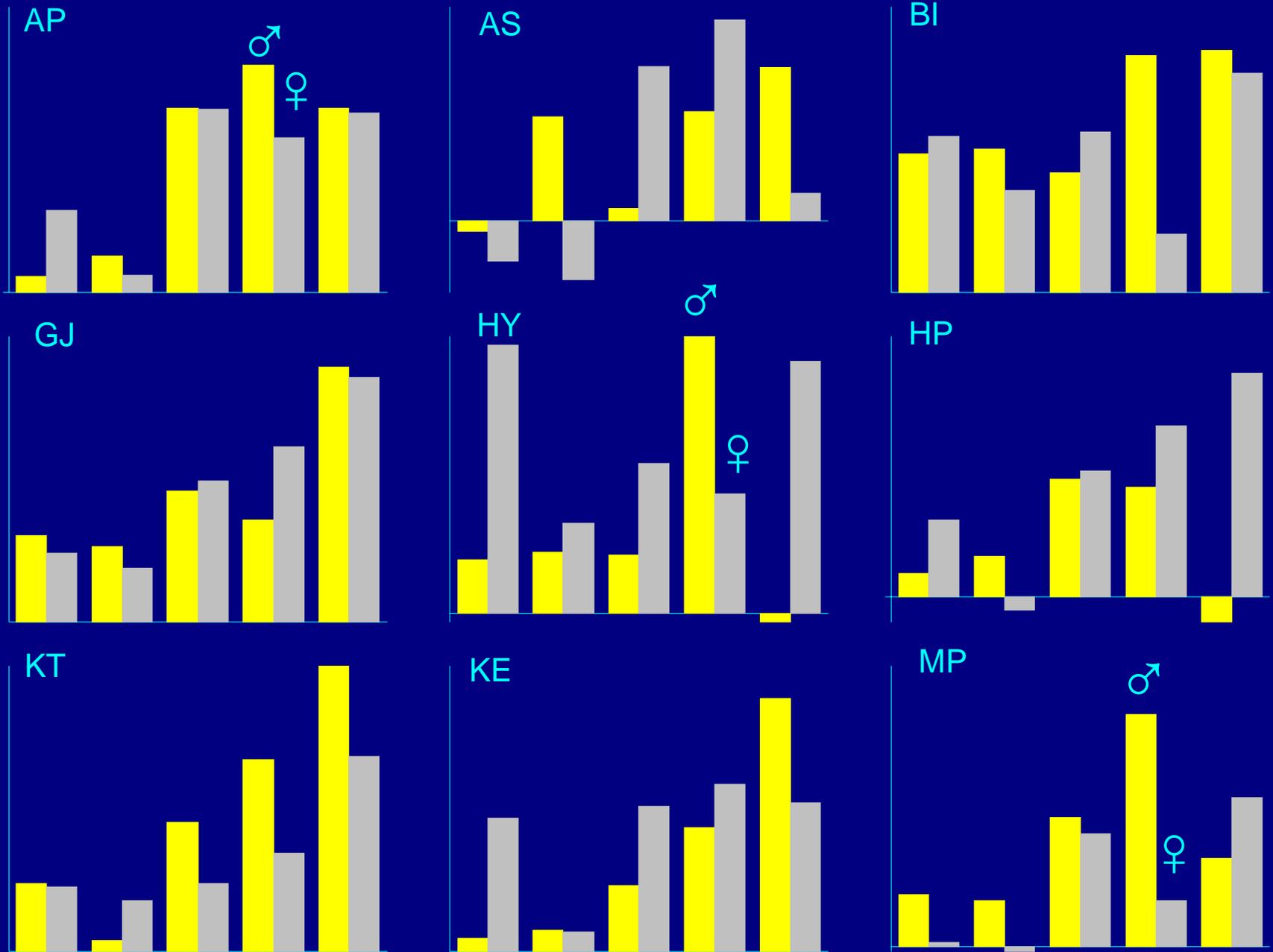


Figure 11: Medical expenditures, rural India

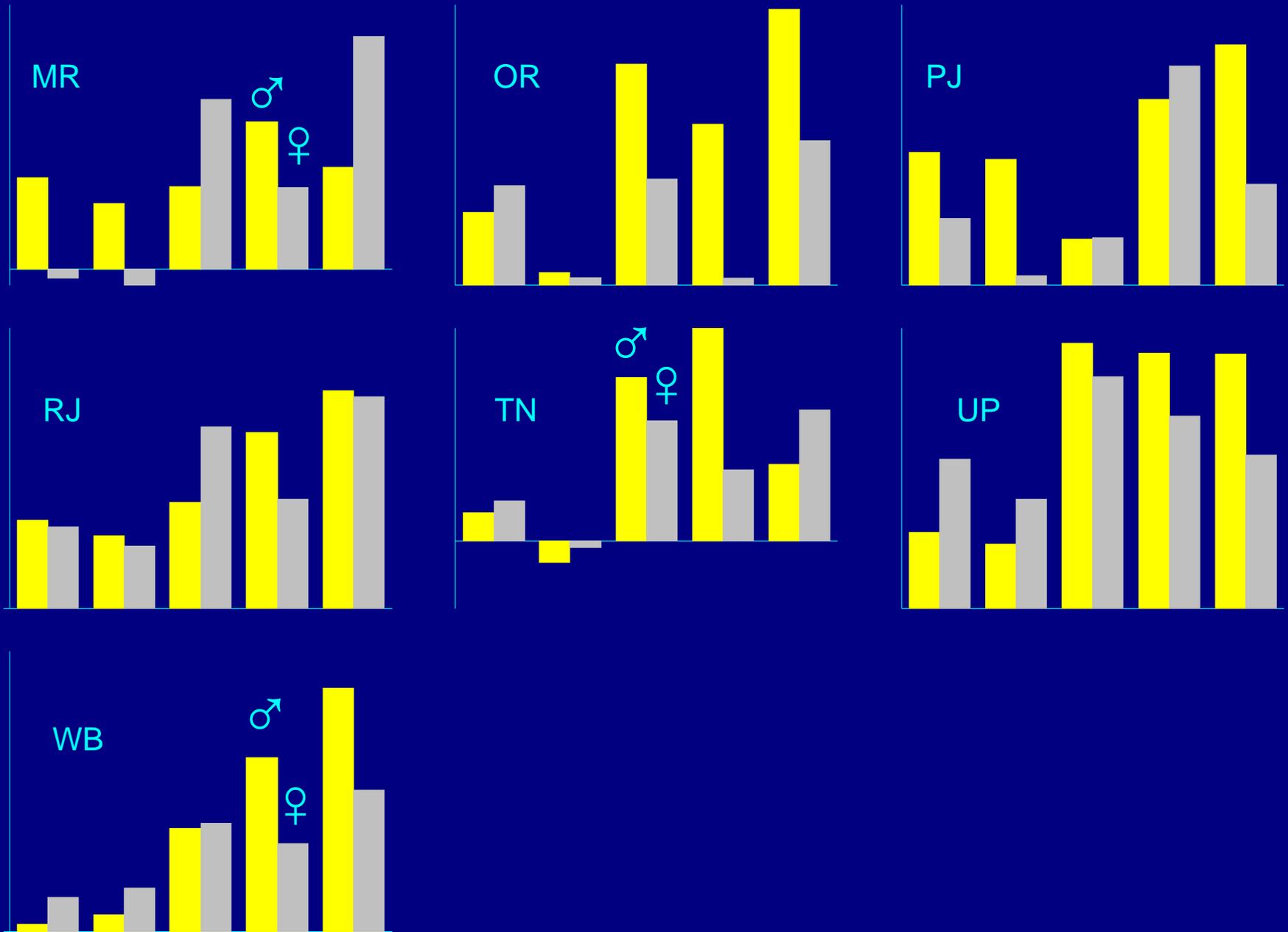
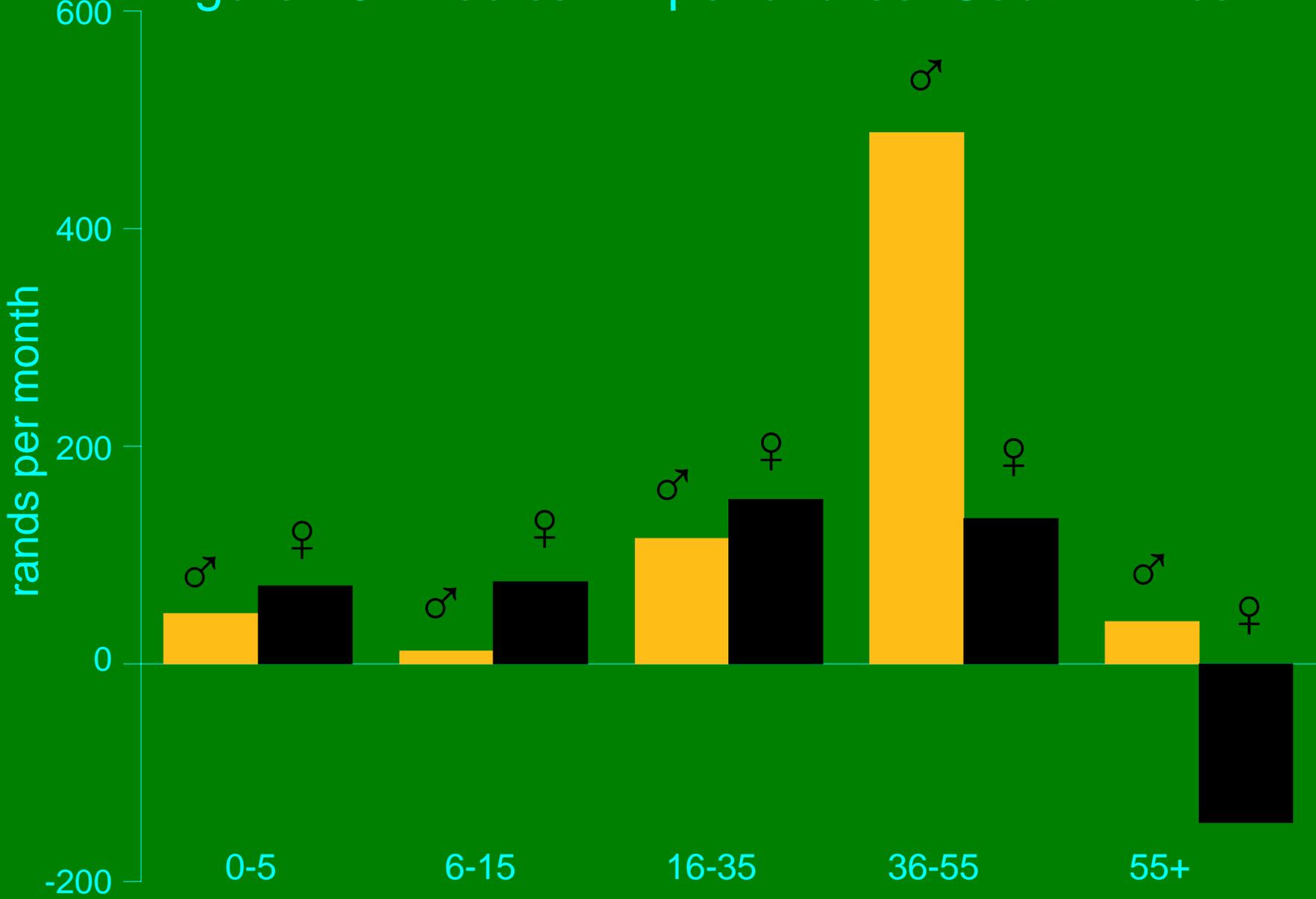


Figure 12: Medical expenditures, rural India

# Figure 13: Medical Expenditures: South Africa



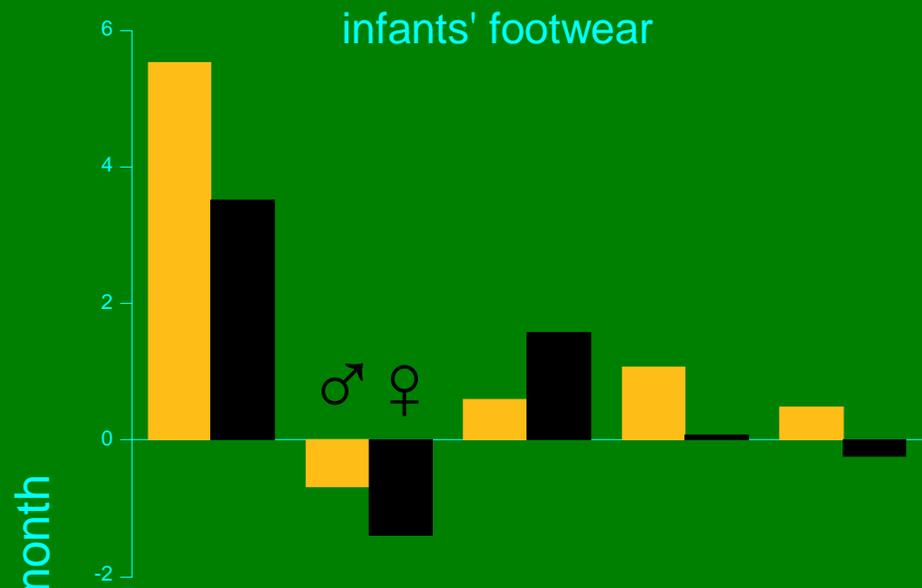
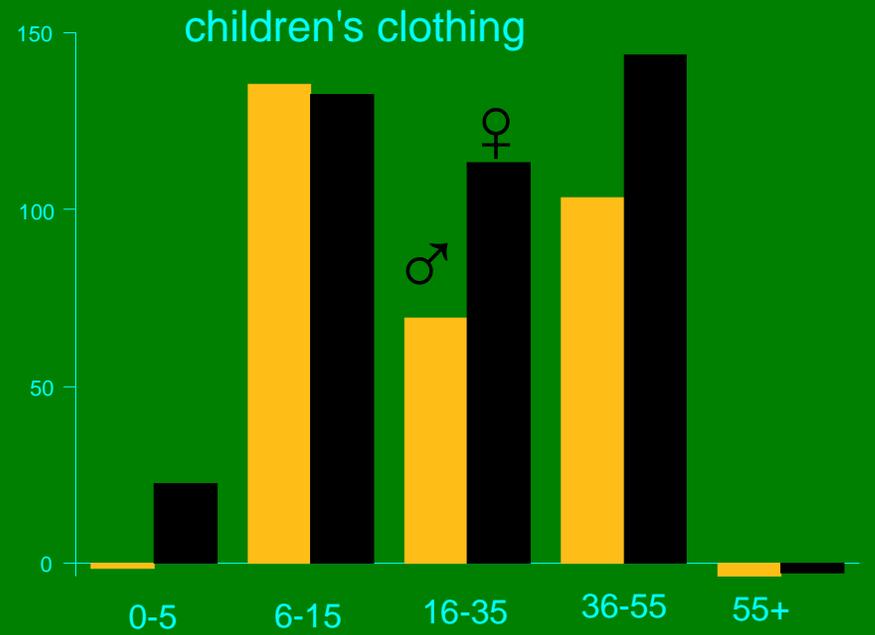
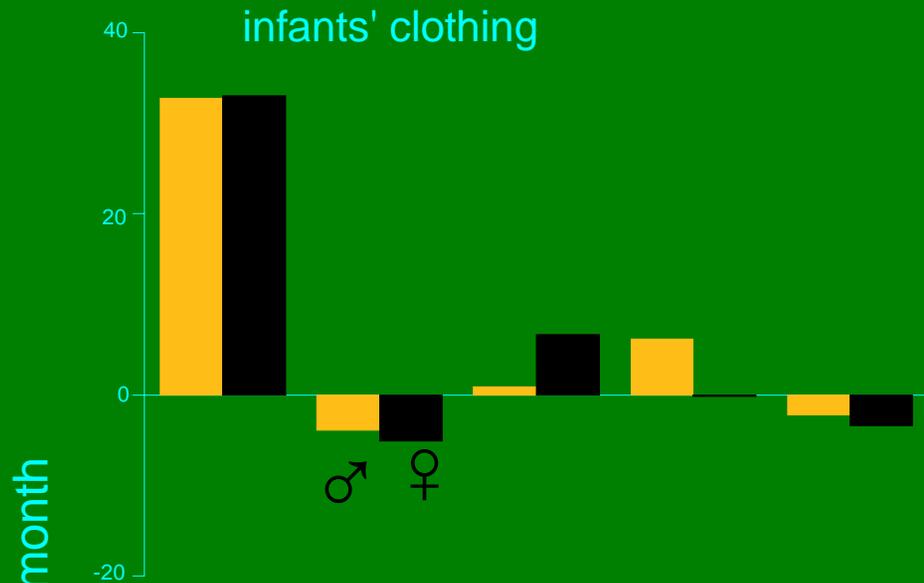


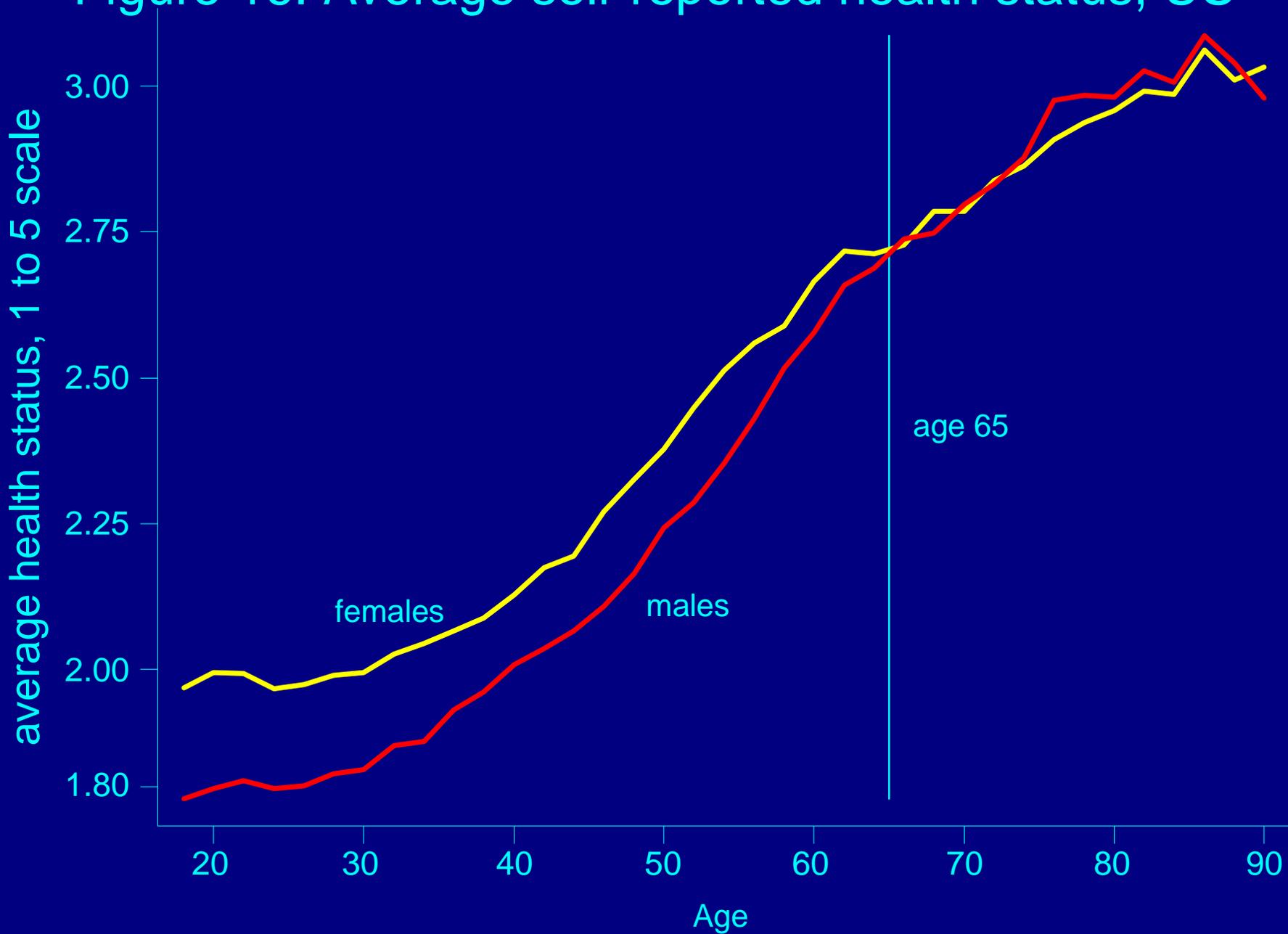
Figure 14: Infant, child and adult footwear



South Africa

Figure 15: Infant, child and adult clothing expenditures

Figure 16: Average self-reported health status, US



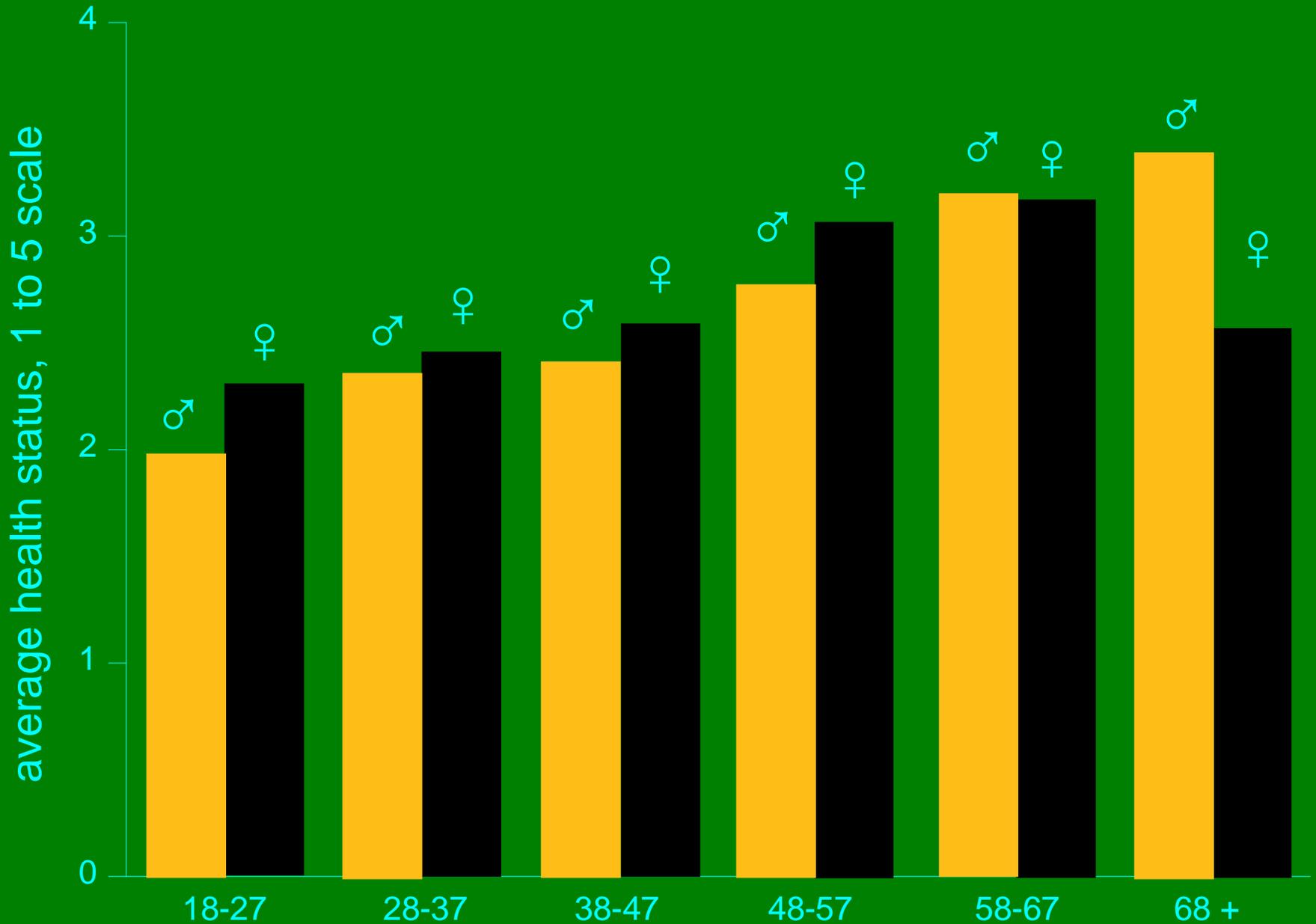


Figure 17: Self-reported health status, South Africa

# FIGURE 18:

Self reported health status by age and quartile of household income, Women and men in the US

