Rationale and tutorial for analysing and reporting sex differences in cardiovascular associations

Mark Woodward [1,2,3]

ABSTRACT

Cardiovascular disease (CVD) is the leading cause of death in women and men. Yet biological and social factors differ between the sexes, while the importance of CVD in women may be underestimated due to the higher age-specific rates in men and the historical bias towards the male model of CVD. Consequently, sex differences in risk factor associations with CVD occur, but these are not always recognised. This article argues that sex disaggregation should be the norm in CVD research, for both humanitarian and clinical reasons. A tutorial on how to design and analyse sex comparisons is provided, including ways of reducing bias and increasing efficiency. This is presented both in the context of analysing individual participant data from a single study and a meta-analysis of sex-specific summary data. Worked examples are provided for both types of research. Fifteen key recommendations are included, which should be considered when undertaking sex comparisons of CVD associations. Paramount among these is the need to estimate sex differences, as ratios of relative risks or differences in risk differences, rather than merely test them for statistical significance. Conversely, when there is no evidence of statistical or clinical significance of a sex difference, the conclusions from the research should not be sex-specific.

RATIONALE FOR STUDYING SEX DIFFERENCES IN CARDIOVASCULAR DISEASE

Historically, cardiovascular disease (CVD) was seen as a disease of men. Nowadays, there is wide recognition that it is also a disease of women, although awareness of coronary heart disease (CHD) as the leading global cause of death in women is lacking both in clinicians and the general public.[1] Indeed, CHD and stroke each appear in the top 4 causes of death in both high-income and low-income countries in women as well as in men.[1 2]

Many authors and activists have done sterling work in raising the profile of female CVD, but often this has involved studies and review articles that report data on women alone. Obviously this is unavoidable if the subject is entirely female—the future risk of CVD in mothers associated with pre-eclampsia during pregnancy, for example—but too often the opportunity of including male controls is missed, leaving open the question of how best to explain the findings from whatever problem the research has investigated. An example comes from a programme of work on women's reproductive factors and the future risk of CVD in the China Kadoorie Biobank. One potential risk exposure addressed was the number of children to

which a woman had given birth.[3] Leaving aside the relatively few women who did not have children, risk increased as the number of children increased. Biological mechanisms that suggest such an association include pregnancy-induced alterations in the cardiometabolic system, which previous authors had suggested.[4] However, unlike the other female reproductive factors—such as breast feeding—analysed in this programme, the relationship between the number of children and CVD risk could also be explored in the men recruited to the Biobank. Male associations were found to be virtually identical, which suggests that the association is not driven by female biology; this changes the interpretation of the results completely.

Even if the main purpose of the research is not to study female issues in CVD, a strong case can be made for systematic inclusion of sex-specific results in the study's report. For example, large-scale clinical trials routinely report subgroup analyses, based on demographic and clinical features of the patient population. Even when one sex predominates for the disease under study, sex should be included as one of the subgroup criteria; that is, the association between the randomised intervention and the study outcome(s) should be reported for both women and men. In most cases this should be one of the prespecified subgroup analyses, which provide the strongest evidence that the finding is not due simply to chance, although allowance must be made for the degree of scientific interest in competing patient characteristics, being mindful of the attenuating impact when many subgroup analyses are specified a priori. While acknowledging the possible drawback of irresponsible overinterpretation of underpowered chance findings,[5] I consider that sex stratification should be reported even when the chance of finding a sex difference is small. This is so that such results can be used in future meta-analyses.

The case for habitual reporting of results for each sex separately is based on three facts. First, women and men are biologically different, and frequently have different social experiences. Exploring sex differences may, thus, uncover important mechanisms, furthering health and medical research. Second, they each make up approximately 50% of the population, so we can expect sex-specific findings, when distinct, to have widespread relevance—no other factor has such a degree of balance, in general. These are obvious statements, and yet sex-disaggregated results are not (yet) commonplace. For example, when conducting a systematic review of sex-specific associations between smoking and stroke,[6] authors found that over 40% of papers,

discovered using keywords that included sex-related terms, failed to present associations for women and men separately. This would not matter if there were no sex differences, but the third arm of the case for sex disaggregation is that there are. Women are less likely to demonstrate the 'classic' symptoms of CVD,[7] which have historically been determined from data on men. Women may thus be misdiagnosed—likely leading to adverse outcomes.[8] More generally, failing to consider the sexes independently can cause incorrect inferences to be concluded, with adverse consequences for one or both sexes.[9 10] Recognition of these issues has led some research bodies and journals to support, or even mandate, the inclusion of, and reporting of results for, both women and men in research studies.[11–14] Although this is far from universal, the contemporary social climate, at least in the Western world, suggests that reporting of results by sex may well become routine in future medical research.

Practical considerations often mean that one cannot balance the sex representation in a study. In an ideal world the same number of women and men would be recruited into all cardiovascular trials, with sex being one of the factors on which randomisation is stratified. But, just as in the case of prespecified subgroups, other factors may trump sex for prognostic value, given that few factors can realistically be allowed for in randomisation. The result may well be a sex imbalance. In epidemiological studies, the sex ratio will typically be dictated by the study population at large, or that of the subpopulation of cases of the index disease. Analyses of sex differences using unequal numbers of women and men are less efficient than otherwise, but this is certainly not a terminal problem.

TUTORIAL FOR ANALYSING AND REPORTING SEX DIFFERENCES IN CARDIOVASCULAR ASSOCIATIONS
Sex and gender

I am taking 'sex' to be the dichotomy between women and men. For behavioural risk factors, 'sex' would usually be replaced by 'gender', but I make no such distinction in this exposition because the methodologies would be the same. Should sex or gender be considered non-binary (as they undoubtedly are), things get more complicated, although once one group is considered to be the reference group generalisations can be made. Similarly, I am restricting myself to examining sex-specific associations between a binary risk factor (eg, obese vs not) and the risk of CVD. The same methods will often apply if the risk factor is considered in a linear continuous way (eg, per $5\,kg/m^2$ body mass index), on the log scale. Should the risk factor be categorical, or at least considered in this fashion (eg, underweight/normal weight/overweight/obese), one group should be considered as the reference (eg, normal weight) and the methods presented here can, again, be generalised by analysing a set of comparisons with the common reference.

Comparison of risk

When studying associations between a risk factor and CVD, the most basic summary statistics are the risks of CVD in the risk factor-positive (exposed) and risk factor-negative (unexposed) subgroups. If the study is prospective with no (or, at least, insignificant) loss to follow-up (censoring), then both risks are relative frequencies (see table 1 for examples). When censoring occurs (due to a death from a non-CVD cause or due to emigration), it is best to allow for it by using the rate, for example per thousand person-years of observation, to estimate the risk.

Risks can be compared on the relative or absolute scale. Comparisons are typically made, respectively, using the relative risk (RR) and risk difference (RD)—see table 1. When one considers that the risk is itself a relative parameter (eg, the number of myocardial infarction (MI) events divided by the number at risk), use of the RR has the advantage of maintaining consistency. RRs and, most tellingly, their variances are also easier than RDs to obtain from standard software, and consequently are much more common in medical literature. Finally, they travel well, in the sense that an RR for a given risk factor—disease association in one study population is likely to be a reasonable estimate of the same association in another population. However, experience shows that typical RRs in CVD research vary with age[15] (eg, see figure 1), so this advantage may be limited to populations with similar age structures. On the other hand, RDs tend to be heterogeneous. They are more useful for making clinical decisions in a specific population, particularly because they can be used to determine the expected number that will experience the outcome studied over a fixed time period. Several eminent epidemiologists have championed their use.[16]

It may well be sensible to analyse the same set of data on both the relative and absolute scales to enable a full set of inferences and interpretations. For instance, in figure 1, when moving from RRs to RDs, not only do the sex-specific patterns in trends by age interchange position (women doing worse for RRs but better for RDs, at all ages) but also so does the direction of trend (RR decreases with age; RD increases).[17]

Here, I am concerned with comparing an association, say between obesity and CVD, between the sexes. This leads to consideration of the ratio of RRs—the RRR—and the difference in RDs—the DRD, as defined in table 1. Since women typically suffer CVD 5–10 years later than men, in all but the most elementary analyses, adjustment by age is essential to obtain meaningful inferences regarding sex differences. We might also like to adjust for other CVD risk factors thought to be confounders. Hence statistical models will be used.

Analysing sex differences using the individual participant data from a single study
Sex comparisons on the relative scale

We can easily analyse the sex-specific portions of individual participant data (IPD) separately to get female and male RRs, which is what many authors do. Having done that, they will often fit the interaction model, for example the model with sex (female/male), obesity (yes/no) and their interaction, to test whether the sex difference is significant. Note that this interaction model should also include all main effects and sex interactions for each confounding variable included in the sex-specific models, otherwise the adjustments made in the interaction model will not vary by sex, as they do in the sex-specific models. However, p values have limited utility, and it is more practically useful to estimate the RRR (and 95% CI) for the sex difference (eg, in the effect of obesity on CVD), stating clearly whether women are compared with men, or vice versa (maintaining consistency, should several similar analyses be conducted in the same study). While the estimated RRR arises from a simple division, deriving its 95% CI from the sex-specific 95% CIs (and estimates) is time-consuming (as described later in the context of a meta-analysis) and inaccurate. Instead, the 95% CI can be found from the computer output when fitting the interaction model; it will usually be given alongside the p value. Note that some computer packages will give results on the logarithmic scale, thus producing ln(RR)s and the ln(RRR), where ln denotes log to the exponential base. Exponential transformations are then required to obtain final results.

Table 1 Fundamental metrics of risk

Metric	Symbol	Example calculation	Interpretation
Example 1: 1000 obese subjects, of whom 75 develop CVD during follow-up; 1000 non-obese subjects, of whom 30 develop CVD during follow-up.			
Risk*		Obese: 75/1000=0.075; non-obese: 30/1000=0.03	75 in a 1000 obese, and 30 in a 1000 non-obese, develop CVD.
Relative risk	RR	0.075/0.03=2.5	The obese have 2.5 times the risk of the non-obese.
Risk difference	RD	0.075−0.03=0.045	Obesity is associated with an additional risk of 45 in a 1000.
Example 2: half (500) of the obese were women, of whom 35 developed CVD; half (500) of the non-obese were women, of whom 10 developed CVD.			
Risk for women		Obese: 35/500=0.07; non-obese: 10/500=0.02	7 in a 100 obese women, and 2 in a 100 non-obese women, develop CVD.
Risk for men		Obese: 40/500=0.08; non-obese: 20/500=0.04	8 in a 100 obese men, and 4 in a 100 non-obese men, develop CVD.
RR for women	RR_{women}	0.07/0.02=3.5	Obese women have 3.5 times the risk of non-obese women.
RR for men	RR_{men}	0.08/0.04=2	Obese men have twice the risk of non-obese men.
RD for women	RD_{women}	0.07−0.02=0.05	Female obesity is associated with an additional risk of 5 in a 100.
RD for men	RD_{men}	0.08−0.04=0.04	Male obesity is associated with an additional risk of 4 in a 100.
Ratio of relative risks, women to men	RRR	3.5/2=1.75	Women have a 75% greater proportional risk increase associated with obesity, compared with men.
Difference of risk differences, women to men	DRD	0.05−0.04=0.01	Women have an additional increased risk of 1 in a 100 associated with obesity, compared with men.

The table includes a simple artificial example of a cohort study assuming no (or ignorable) censoring during a fixed duration of follow-up. In example 1, the sex of the subject is ignored; in example 2, sex differences are evaluated.

*Often called the 'absolute risk'. I feel that the qualifier is unnecessary and inappropriate because it suggests some kind of truth, whereas in general the risk is merely an estimate subject to random error and, sometimes, bias error. It can also be confusing, as when absolute risk is incorrectly represented as an alternative to the RR (the true alternative to the RR is the RD).

Notes

1. With a cross-sectional study, replace 'risk' by 'prevalence'.

2. With a case–control study, replace RR by 'odds ratio' (OR).

3. With a cohort study that is analysed using logistic regression, replace RR by OR. Censoring is ignored. Often the OR is assumed to be the same as the RR, which is reasonable if the disease analysed is rare in the study population, but the OR will always overestimate the RR. In example 1, the OR is (75/(1000−75))/(30/(1000−30))=2.62, slightly higher than the RR of 2.5.

4. With a cohort study that is analysed using log-binomial regression, risks and RRs are estimated. Censoring is ignored.

5. With a cohort study that is analysed using Cox or Weibull proportional hazards regression models, HRs are estimated. These are generally taken to be the same as the RR. Censoring is accounted for.

6. With a cohort study that is analysed using Poisson models, rates and relative rates are estimated. These are generally taken to be the same as risks and RRs. Censoring is accounted for.

CVD, cardiovascular disease.

An alternative approach is to use the interaction model to obtain the RRs as well as the RRR. This works because the interaction model will directly produce the RR for whichever sex is taken as the reference group, as well as the RRR comparing the other sex with the reference sex. So, if men are taken as the reference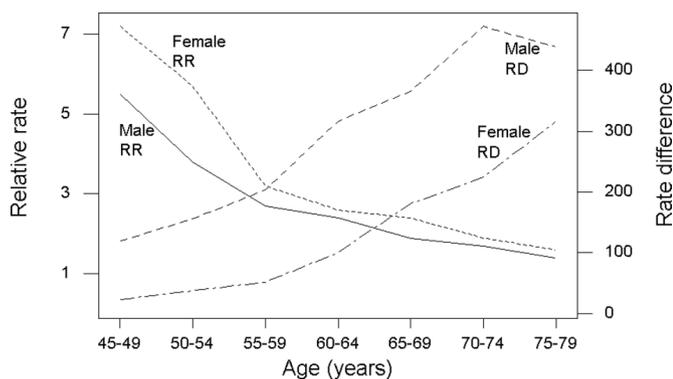 group, we would get the male RR and the female to male RRR (as well as their 95% CIs) straight from the computer output. To get the female RR from the same computer output is quite easy, but getting its 95% CI is less straightforward since covariances must be dealt with. A simple trick, to avoid this issue, is to interchange the codes given to the sex variable (often 0 and 1), thus making the other sex now the reference group, and run the interaction model again. Continuing the example, this second run would produce the female RR and the male to female RRR (and 95% CI). We are unlikely to want the latter, since it is simply the reciprocal of the female to male RRR (the same is true of the confidence limits, once they are interchanged).

I prefer to use the interaction model to obtain the RRs and RRR because it requires fewer models to be fit. Also, using the interaction model alone ensures that results for the ratio of sex-specific RRs and the RRR must be completely concordant.

Figure 1 Relative rate (RR) and rate difference (RD) (per 100 000 per year) for coronary heart disease by age group (45–79 years old) and sex, comparing smokers to never-smokers (American Cancer Prevention Study II, National Cancer Institute, 1997). Figure reproduced from Woodward (*Epidemiology: Study Design and Data Analysis*, Third Edition[17] and Second Edition, 2005).

Worked example

Consider the problem of estimating the RRs by sex and the female to male RRR when relating diabetes to the risk of MI, adjusting for age, systolic blood pressure and smoking status. Using data from the UK Biobank, I selected all people without CVD at baseline, except those (relatively few) with missing values, and fitted Cox proportional hazards regression models in the Stata package (V.15); the code and results appear in online supplementary appendix 1. Using sex-specific models, the RRs for MI (diabetes vs not) were 2.97 (95% CI 2.53 to 3.48) for women

and 1.66 (95% CI 1.49 to 1.84) for men over a 7-year follow-up period. The female to male RRR was thus 2.97/1.66=1.79. The interaction model (with men as the reference group) gave the RR for men as 1.66 (95% CI 1.49 to 1.84) and the female to male RRR as 1.79 (95% CI 1.48 to 2.17). Swapping to use of women as the reference group gave the RR for women as 2.97 (95% CI 2.53 to 3.48). So both methods give the same results, but use of the interaction model alone (here fitted twice) saved one model fitting exercise compared with the sex-specific approach (which needs one run of the interaction model to get the 95% CI for the RRR). In Stata, and some other software, it is possible to get all the three statistics of interest, with CIs, directly by fitting the interaction model just once (see online supplementary appendix 1). Whatever the approach, we conclude that diabetes, after adjustment, increases the risk of MI in both sexes, but the relative increase is about 80% bigger in women. Note that because Cox models were used, strictly speaking, the results here are HRs and their quotient (see notes to table 1).

Sex comparisons on the absolute scale

On the absolute scale, when survival data are analysed, it is usual to use rates per person-year to express risk. Sex-specific RDs and the DRD derive from Poisson regression models, taking the logarithm of the follow-up time (per individual), suitably scaled (so that results, for example, per 10 000 person-years, are produced) as an offset.[17] Fitting a Poisson model with sex as the only input variable gives the sex-specific rates in the study population—although these are easy to compute by hand, Poisson regression also gives the accompanying CIs.

Worked example

Using the same UK Biobank data as before, online supplementary appendix 2 shows the results from fitting Poisson regression models in Stata. The rates of MI were found to be 7.90 and 24.64 per 10 000 person-years for women and men, respectively. This puts the risk of MI in perspective—for every thousand women followed up for 10 years, we would expect about 8 to get MI, and for every thousand men we would expect about 25. Fitting models for sex differences relating diabetes to MI produced estimates of rates (and their CIs; approximately evaluated using the delta method in Stata). Table 2 shows both the unadjusted and adjusted rates, using the same adjustments as when RRs were estimated previously. This shows that women, and those without diabetes, have the lower rates, and adjustment only has any appreciable effect for men with diabetes—the group at most risk. Simple subtractions show that diabetes is associated with roughly one extra MI case per 10 000 person-years of observation in women compared with men. Despite the much higher RR in women found earlier, the difference in additional expected MI cases associated with diabetes is negligible in this study population (without previous CVD and otherwise relatively healthy), partially because the overall risk of MI is low. Should the RRR

Table 3 Multiple-adjusted coronary heart disease relative risks (and 95% CIs) for women and men, comparing those with, to those without, diabetes, by study

Study	Women	Men
Adventist	2.15 (1.33 to 3.47)	2.11 (1.12 to 4.00)
APCSC (ANZ)	2.01 (1.55 to 2.60)	1.58 (1.32 to 1.90)
APCSC (Asia)	1.82 (1.02 to 3.25)	1.47 (1.15 to 1.88)
ARIC	3.16 (2.64 to 3.78)	2.38 (2.02 to 2.80)
Collins (Indians)	20.70 (2.51 to 171)	3.15 (1.29 to 7.69)
Collins (Melanesians)	5.36 (1.18 to 24.3)	1.60 (0.43 to 5.97)
DECODE	2.48 (1.69 to 3.65)	2.09 (1.55 to 2.82)
Dubbo	1.67 (1.12 to 2.48)	1.53 (0.99 to 2.37)
EPESE	3.20 (1.46 to 7.01)	1.75 (0.97 to 3.16)
Framingham	5.4 (2.4 to 12.3)	6.1 (3.4 to 10.9)
Hawaii/LA/Hisoshima	3.29 (1.79 to 6.55)	1.54 (1.03 to 2.25)
Hisayama	3.46 (1.59 to 7.54)	1.26 (0.67 to 2.35)
HUNT I	2.50 (2.10 to 2.80)	1.80 (1.60 to 2.10)
Kuopio and N Karelia	4.89 (3.84 to 6.24)	2.11 (1.70 to 2.63)
NHANES I	2.59 (1.59 to 4.22)	2.37 (1.55 to 3.62)
NHANES III	2.53 (1.62 to 3.97)	1.29 (0.91 to 1.85)
Renfrew and Paisley	1.97 (1.27 to 3.08)	1.17 (0.78 to 1.74)
Reykjavik	2.23 (1.50 to 3.32)	1.34 (0.97 to 1.87)
SHHEC	3.06 (2.18 to 4.27)	2.49 (1.84 to 3.37)
Strong	2.26 (1.73 to 2.96)	1.66 (1.30 to 2.12)
Takayama	0.49 (0.07 to 3.57)	2.96 (1.59 to 5.50)

For citations to the studies (identified here by authors or study names), see Peters et al,[18] from where the data were obtained.

stay the same in high-risk subpopulations, we would expect much higher numbers of excess female cases of MI.

See Millett et al[15] for a comprehensive study of sex differences in risk factors for MI in the UK Biobank using the same methods described here.

Incidentally, the Poisson model can also be used to estimate the RRs and RRR, as an alternative to the Cox model. Comparison of fitted parameter estimates for the adjusted models in online supplementary appendices 1 and 2 shows very similar estimates and 95% CIs. A simple way of seeing the linkage between the two approaches is to compute the adjusted RRs and the RRR from the results in the right-hand side of table 2, which gives the RRs as 2.97 (women) and 1.66 (men), with the RRR of 1.79, just as was found from the Cox model.

Analysing sex differences using summary data from multiple studies

A number of different studies may have published sex-disaggregated results linking the risk factor of our interest to the outcome of our interest. It would then make sense to summarise their findings using meta-analysis, starting from a systematic review of the literature (possibly including unpublished sources and certainly including citation tracing from original sources). Suppose that every study has an estimate of the RR (per sex), and these are what we wish to summarise, through pooling. As a motivating example, I will take the study of Peters et al.[18] The authors compiled data from 19 studies which provided sex-specific RRs for CHD, comparing those with diabetes to those not; two of these studies provided RRs in two separate subgroups of their total study population, giving 21 RRs to pool in all, for each sex (table 3).

Since our goal is to compare the sexes, it is best to only include studies with results for both women and men. This is because differences in the make-up of study populations, definitions of the
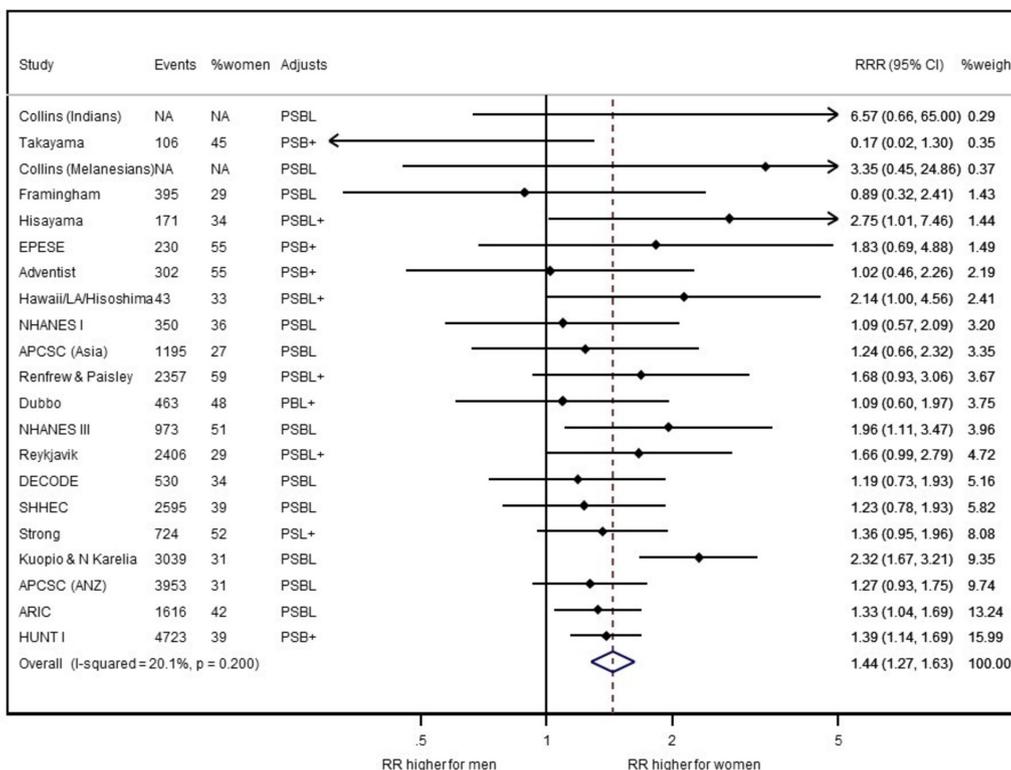
Table 2 Rates of myocardial infarction, per 10 000 person-years, in a subsample of the UK Biobank without cardiovascular disease at baseline

	Unadjusted analysis		Adjusted* analysis	
	Women	Men	Women	Men
Diabetes	25.74	45.54	23.71	36.66
No diabetes	7.25	23.31	7.99	22.13

*Adjusted for age, systolic blood pressure and smoking.

Figure 2 Women to men ratios of coronary heart disease relative risks (RRRs), comparing those with, to those without, diabetes, by study and pooled overall. Data are from table 3. Random effects inverse variance weighting was used to pool the study-specific data. Horizontal lines show 95% CIs, as does the width of the summary diamond. 'Events' are of coronary heart disease during follow-up (some studies only recorded fatal coronary events), and '%women' gives the percentage of these events that were female. 'NA' denotes 'not available'. 'Adjusts' gives the summary details of the adjustments made, per study: P denotes blood pressure (which is most often systolic, but sometimes is hypertension or antihypertensive use; in one study adjustments were made for both diastolic and systolic blood pressure); S denotes smoking; B denotes body mass index; L denotes lipids (which always included total cholesterol, but sometimes also other lipids); + denotes other coronary risk factors. The p value is for a test of no heterogeneity.[17] RR, relative risk; RRR, ratio of RR.

exposure and outcome, and the methods of analysis might introduce bias error if single-sex studies are included in the pooling process. For instance, published studies must be expected to include a range of adjustment sets; the data in table 3 come from studies which adjusted for confounding factors in many ways, with between 5 and 10 factors included (see figure 2). By only including studies with results from both sexes, we ensure that between-study variations (known or unknown) will be the same for women and men. Another general exclusion, which I would advise, is studies that do not adjust at least for age. Often other classical risk factors for CHD will also be adjusted for in published studies, but being too prescriptive in what must be included or excluded can lead to few studies being selected for pooling. If we have access to IPD, we can decide on our own adjustments. Indeed, four of the RR pairs in table 3 came from IPD analyses, adjusting diabetes for age, systolic blood pressure, smoking, body mass index and serum total cholesterol in each case.

Study pooling

To obtain an overall picture of the sex-specific association between the exposure and the disease (diabetes and CHD, in the example), meta-analysis can be used to estimate the pooled RR, across all studies, for women and for men. There are many ways that such pooling can be achieved.[17] For example, we could take a simple mean of the separate study estimates, but this would not account for any differences in quality (relative lack of bias) or quantity (precision; relative lack of sampling error) between the constituent studies. So a weighted mean is preferred. In practice,

most meta-analyses weight by precision and leave bias to be assessed more informally using standard criteria, such as the Newcastle-Ottawa Scale,[19] independent of the pooling process.

The most popular weighting scheme is inverse variance (IV) weighting—as the name suggests, each study is weighted by the reciprocal of its variance. This makes sense because the less precise the study, the larger will be its variance (and the wider will be the CI around the point estimate of the RR). Subsequently all weights are transformed to sum to 100%, for easy interpretation, by dividing each individual weight by the sum of all the weights and multiplying by 100. An advantage of IV weighting is that it produces the narrowest possible CI for the pooled RR. A drawback, shared by other popular weighting schemes, is that it can only be applied to studies for which some measure of the variability associated with the study's estimate is available. Generally published studies report the SE of the estimate (ie, the square root of the variance) or a 95% CI. To pool results we need the same metric for all studies; suppose we decide to use the SEs. We may then need to derive SEs from published 95% CIs in some cases. This requires, first, taking logarithms of the estimate and the 95% confidence limits. This is necessary because, although the RR itself does not have the classic normal form, its logarithm, ln(RR), is approximately normally distributed. That being so, we know that the 95% CI must have the form[17]:

$$(E - 1.96SE, \; E + 1.96SE)$$

Table 4	Worked examples using the first study in table 3 (the Adventist study)
1	The female RR (and 95% CI) is 2.15 (1.33 to 3.47). Taking logs of all three numbers gives the ln(RR) and its 95% CI: 0.765468 (0.285179 to 1.244155). The two equations for SE give the results 0.244228 and 0.245045, which average out to 0.244637, our best estimate for the SE of the ln(RR).
2	Similar computations for men give corresponding results for the ln(RR) and SE of 0.746688 and 0.324736. The ln(RRR) is thus 0.765468−0.746688=0.01878, and its variance is $0.244637^2 + 0.324736^2 = 0.165301$. The 95% CI for the ln(RRR) is $0.01878 \pm 1.96\sqrt{0.165301} = (-0.778101$ to $0.815661)$. The estimated RRR and 95% CI are then the exponents of the ln(RRR) and its confidence limits, that is, 1.02 (0.46 to 2.26), to two decimal places.

RR, relative risk; RRR, ratio of RRs.

where E is the point estimate, ln(RR), and SE is the SE of the ln(RR). This means that

$$SE = \frac{(E-L)}{1.96} \text{ and } SE = \frac{(H-E)}{1.96}$$

where L and H are the lower and higher 95% confidence limits, respectively, of the ln(RR). In practice, these two equations will give different answers for the SE, if only due to rounding errors in L and H. If they are very different we should check our arithmetic and refer back to the source document for confirmation (numerical errors in published works have been known to occur). The two results should be averaged to find the 'best' estimate of the SE (see table 4, example 1).

Once we have estimates, and their variability, these can easily be pooled. The computations are not complex for IV weighting[17] and could be applied in Excel. However, most researchers will want to use software, not least to be able to produce graphical displays. A range of specialist software exists,[20] but I prefer to use the user-supplied Stata procedure *metan*. This procedure allows the user to enter either the estimate and its SE or the estimate and its 95% CI (or a CI of a different degree) from each study to commence pooling and assumes that the estimate follows a (near) normal distribution, at least for the type of data addressed here. As already discussed, the RR does not follow this form, so when meta-analysing RRs we should instead pool ln(RR)s. The *eform* option should be included in the *metan* command as this back-transforms the natural logs of the pooled RR, and its confidence limits, to the natural scale, in the presentation of results.

Fixed effect and random effects

This account of meta-analysis, so far, has implicitly assumed the fixed effect approach. This assumes that there is a universal true RR for women (and for men) which relates to all studies, wherever, or whenever, they were conducted. The variation between estimates is attributed solely to study-specific sampling error—that is, the difference between the sample and its source population. Many, but not all, authors prefer the random effects approach which allows the true RR to vary by study and interprets the pooled estimate as the average of all the different 'truths'.[17] Random effects allows for greater uncertainty and thus will produce 95% CIs for the pooled RR that are at least as big as for the corresponding fixed effect analysis. To the practitioner, the advantage of random effects is that it will give the same result as fixed effect if there is no between-study heterogeneity, so in this sense it is the safe option. It operates like IV

weighting, as it has been described above, but where an additional component is added to the variance for each study to account for between-study heterogeneity. That is, fixed effect weights a study by 1/V, where V is that study's variance, but random effects weights by 1/(U+V), where U is the between-study variance (computed from a complex formula[17 21]; omitted here). As a consequence, the random effects approach gives less weight to the biggest (in terms of outcome events) cohort studies than does the fixed effect.

Putting the data from table 3 into Stata, using the *metan* procedure to pool the RRs for each sex, produced the results in table 5. As can be seen, similar results pertained whether fixed effect or random effects was used. It is also very clear that the RR for women is larger than that for men, suggesting that diabetes is a stronger risk factor for CHD in women than in men. An estimate of the percentage of variability attributable to between-study heterogeneity, called the *i-squared* statistic,[17 21] has been included in table 5. Some researchers use this to decide whether to apply fixed effect or random effects, often using 25% as a threshold above which between-study heterogeneity is thought to be too great for the fixed effect assumption to be viable. Others, including me, feel that the choice should be made a priori. As in all cases, methods decided on before the data are analysed will be more defensible.

The ratio of relative risks

So far, no mention has been made of sex comparisons, which can be made through the women to men RRR. Given the female and male RRs, the RRR is straightforward to derive per study, but to pool the RRRs using IV weighting requires knowledge of the variance (or similar) of the RRR in each study. These are simple to estimate once we transform to the log scale, which converts the problem from dealing with ratios to one dealing with differences. The variance of the difference in ln(RR)s between women and men is the sum of the variances of the female and male ln(RR)s, and the SE of the difference is its square root. Since this difference can reasonably be assumed to follow a normal distribution, we can compute the 95% CI for the ln(RRR), if required, using the formulae given above, and use *metan* to pool the ln(RRR)s across studies. Table 4 (example 2) has the calculation for the first study in table 3, and online supplementary appendix 3 is a copy of the Excel file used to make computations for all 21 studies (or part studies).

Table 5	Inverse variance weighted pooled relative risks and ratios of relative risks (with 95% CIs) for the association between diabetes and coronary heart disease			

Meta-analysis method	Relative risk		Ratio of relative risks Women:Men
	Women	Men	
Fixed effect	2.68 (2.49 to 2.89)	1.85 (1.74 to 1.97)	1.43 (1.30 to 1.58)
Random effects	2.63 (2.27 to 3.06)	1.85 (1.64 to 2.10)	1.44 (1.27 to 1.63)
	(i^2=64.7%)	(i^2=66.0%)	(i^2=20.1%)

Data from table 3 were analysed.

Woodward M. *Heart* 2019;**105**:1701–1708. doi:10.1136/heartjnl-2019-315299

Pooling of RRRs across studies then proceeds in the same way as for the RRs, that is, pooling on the log scale and back-transforming the pooled estimate and 95% confidence limits. Figure 2 is a forest plot[17 21] of the individual study and random effects IV weighted pooled results. To produce this plot I used *metan* in Stata, having input the 21 ln(RRR)s, and their SEs, from online supplementary appendix 3, augmented by some additional study characteristics for enhanced interpretation, as presented in online supplementary appendix 4. The Stata code and results appear in online supplementary appendix 5, while table 5 includes summary results from both fixed effect and random effects analyses. I did minor artistic editing of the Stata graph in PowerPoint.

Forest plots are extremely useful for presenting the results of a meta-analysis, as well as in other settings; for example, Millett *et al*[15] used them to compare sex differences in MI across a range of risk factors. Rather than simply order the lines in the plot alphabetically (as would be common in a list of studies), more information is gleamed if they are ordered by study weight, as in figure 2, or according to some important feature of the data, such as year of study publication. However, when female and male results are shown alongside each other, a common ordering aids interpretation. Peters *et al*[18] included a forest plot ordered by the size of the estimated RRR, which is also informative and makes for, arguably, the most visually attractive presentation. These authors[18] also included a box around each study's estimate, with boxes drawn in proportion to the precision of the study (ie, the IV). This offers an additional visual idea of the relative weight contributed by each study to the pooled estimate.

From figure 2, we conclude that, although diabetes increases the risk of CHD in both sexes, women have a 44% higher RR for CHD related to diabetes than men. We are 95% confident that the range from 27% to 63% contains the true excess RR. Since this interval omits 0% by a considerable degree, there is statistical evidence of a real sex difference.

Notice that when the female RR given in table 5 is divided by the corresponding male RR, the result is not quite the same as the RRR produced by pooling. This is not unexpected, but is rarely an issue since the two tend to be very similar, as here.

Other considerations in report writing

In manuscripts, a flow chart describing the systematic review that identified the data used in the meta-analysis is essential. I would also recommend showing female-specific and male-specific RRs, as well as RRRs, in forest plots. This is both because the sex-specific results are themselves of interest and because it makes the RRRs easier to understand. Comparing age-adjusted and multiple-adjusted results may provide another insightful contrast.

As with all meta-analyses, investigation of the causes of heterogeneity is essential through subgroup analyses and meta-regression, with accompanying bubble plots.[17 22–24] Typical factors to investigate as causes of heterogeneity in the RRRs are age, year(s) of study, length of follow-up, risks (overall average or their female to male difference) and the prevalence of the index risk factor (again, overall or as a sex contrast), but what is possible is driven by the published data and thus hard to decide completely a priori. It is also useful to identify any influential studies and investigate possible publication bias, taking remedial action where appropriate.[17 22]

Other metrics of sex difference

In cases where ORs, HRs or relative rates are to be pooled, the same methods as above can be used to obtain the sex-specific results and the sex comparison (eg, through the ratio of ORs). Indeed, meta-analyses often pool published results claimed to be RRs from studies which have used different metrics to measure relative associations. In most cases this is very reasonable, but care is required if ORs are mixed with other metrics, because

Table 6	Recommendations
General	
G1	Consider whether the research is concerned with sex (biological) or gender (behavioural) differences, and report the results accordingly[*].
G2	Routinely provide sex-disaggregated results when reporting research on cardiovascular associations. This includes prespecifying subgroup analyses by sex. When there are no important sex differences, still include sex-specific results, most likely in the appendix of a manuscript for publication.
G3	Even when a study is concerned with associations for a single sex, where possible compare results for the other sex, as a control.
G4	Adjust at least for age when comparing sex-specific cardiovascular associations.
G5	Consider analyses on both the relative and absolute scales. When it is only appropriate to present relative risks, provide (at least) the number of events and the number at risk across the sex by risk factor exposure cross-classes, to give context to the reader.
G6	Quantify the sex difference (with accompanying measure of uncertainty, such as a 95% CI), rather than merely test for a significant difference.
G7	When analysing raw (ie, individual participant) data, use the full interaction model (with all main effects and two-way interactions) to obtain the sex-specific results, as well as the sex comparison(s).
G8	Unless there is statistical or clinical significance in the sex difference (ie, the sex interaction), avoid sex-specific conclusions.
Specific to meta-analyses	
M1	Decide whether to use the fixed effect or random effects method before data are collected.
M2	Only include studies with results from both sexes.
M3	In the report, include a flow chart with reasons for exclusions. Clearly state the number of studies excluded for want of sex-disaggregated results.
M4	Use reliable, general, statistical software[†], such as R or Stata.
M5	Include forest plots by sex and to compare the sexes[‡]. Show age-adjusted and multiple-adjusted analyses separately, where appropriate. This will typically require placing some forest plots in the appendix of a manuscript for publication.
M6	Following the meta-analysis, use meta-regression and bubble plots to explore sources of heterogeneity, to include overall risk and the difference between the sex-specific risks.
M7	Take care when pooling ORs together with relative risks or HRs. Stratify pooling by the metric used where risk (or, in cross-sectional studies, prevalence) is typically high.

*In this manuscript no distinction is made, for simplicity of exposition.
†These have the advantage of offering a wide range of other tools, so that the extra work of learning the basics of such a package (if necessary) will be worthwhile.
‡For example, through the ratio of relative risks—see figure 2.

the OR overestimates the RR. The general rule is that the OR is an acceptable proxy for the RR if the outcome analysed is rare, which is the case for CVD in most general populations. However, this is usually not the case when comparing the prevalence of anti-CVD medication use between the sexes in secondary prevention. Better, then, to show pooled results separately for each metric used.[25]

When analyses and inferences on the absolute scale are envisaged, the RD would be the measure of association in each sex, and the sex effect would be estimated by the DRD. Unfortunately pooling is rarely justified in this case because risk tends to be extremely variable between study populations and over time. For instance, while one might reasonably hypothesise that the RRs, comparing those with and without diabetes, for CHD are similar across the world, it is much more of a stretch to think that the rate of CHD per 10 000 person-years in those with (or without) diabetes is very similar in relatively poor and relatively rich countries, or the same 50 years ago as now in the same country. In the vast majority of cases there is too much heterogeneity to make pooled estimates of risk, or RD, meaningful (no average could have a useful meaning). Another problem is that most published studies do not provide the necessary statistics for meta-analyses; they may not even state the sex-specific risks with and without the risk factor of interest, and associated variability. Sometimes the only way of bringing in issues of absolute effects is to follow a meta-analysis on the RRRs with meta-regression of the RRRs on the CVD risks, or their differences by sex, for those studies which provide them, to decide whether the pooled RRR can reasonably be assumed to represent most scenarios.

CONCLUSIONS

Sex disaggregation of research findings is strongly encouraged. Methods employed to investigate, and compare, sex-specific associations are generally straightforward, but interpretation can require some careful thinking. Unveiling a sex difference naturally leads to investigation of its cause, which is a natural extension of the subject matter described in this manuscript. For example, the excess proportionate risk of heart disease associated with diabetes in women, compared with men, found here in two distinct research databases, leads to further research into the cause.[26]

A summary of my recommendations appears in table 6. This may serve as a useful checklist when undertaking sex differences research in CVD. Stata and Excel files used in my analyses are given in the online supplementary material.

Data sharing statement Researchers can apply to use the UK Biobank resource and access the data used. No additional data are available.

ORCID iD
Mark Woodward http://orcid.org/0000-0001-9800-5296

REFERENCES

1 Woodward M. Cardiovascular disease and the female disadvantage. *Int J Environ Res Public Health* 2019;16:1165.
2 Institute for Health Metrics and Evaluation. Global burden of disease study. https://vizhub.healthdata.org/gbd-compare/ (Accessed 6 Feb 2019).
3 Peters SA, Yang L, Guo Y, *et al*. Parenthood and the risk of cardiovascular diseases among 0.5 million men and women: findings from the China Kadoorie Biobank. *Int J Epidemiol* 2017;46:180–9.
4 Williams D. Pregnancy: a stress test for life. *Curr Opin Obstet Gynecol* 2003;15:465–71.
5 Bailey KR. Reporting of sex-specific results: a statistician's perspective. *Mayo Clin Proc* 2007;82:158.
6 Peters SA, Huxley RR, Woodward M. Smoking as a risk factor for stroke in women compared with men: a systematic review and meta-analysis of 81 cohorts, including 3,980,359 individuals and 42,401 strokes. *Stroke* 2013;44:2821–8.
7 Aggarwal NR, Patel HN, Mehta LS, *et al*. Sex differences in ischemic heart disease: advances, obstacles, and next steps. *Circ Cardiovasc Qual Outcomes* 2018;11:e004437.
8 Wu J, Gale CP, Hall M, *et al*. Editor's Choice - Impact of initial hospital diagnosis on mortality for acute myocardial infarction: A national cohort study. *Eur Heart J Acute Cardiovasc Care* 2018;7:139–48.
9 Wizemann TM, Pardue M. *National Research Council, Institute of Medicine Committee on understanding the biology of sex and gender differences. Exploring the biological contributions to human health: does sex matter?* Washington, DC: The National Academies Press, 2001.
10 de Vries ST, Denig P, Ekhart C, *et al*. Sex differences in adverse drug reactions reported to the National Pharmacovigilance Centre in the Netherlands: an explorative observational study. *Br J Clin Pharmacol* 2019;85.
11 Tannenbaum C. Institute of Gender and Health Montreal: Canadian Institute of Health Research. 2013 http://www.cihr-irsc.gc.ca/e/8673.html (Accessed 26 Apr 2019).
12 Clayton JA. Studying both sexes: a guiding principle for biomedicine. *Faseb J* 2016;30:519–24.
13 Schiebinger L, Leopold SS, Miller VM. Editorial policies for sex and gender analysis. *Lancet* 2016;388:2841–2.
14 Heidari S, Babor TF, De Castro P, *et al*. Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. *Res Integr Peer Rev* 2016;1:2.
15 Millett ERC, Peters SAE, Woodward M. Sex differences in risk factors for myocardial infarction: cohort study of UK Biobank participants. *BMJ* 2018;363:k4247.
16 Poole C. A history of the population attributable fraction and related measures. *Ann Epidemiol* 2015;25:147–54.
17 Woodward M. *Epidemiology: Study Design and Data Analysis*. 3rd edn. Boca Raton: CRC Press, 2014. ISBN-13: 978-1-4398-3970-6.
18 Peters SA, Huxley RR, Woodward M. Diabetes as risk factor for incident coronary heart disease in women compared with men: a systematic review and meta-analysis of 64 cohorts including 858,507 individuals and 28,203 coronary events. *Diabetologia* 2014;57:1542–51.
19 Wells G, Shea B, O'Connell D, *et al*. The Newcastle–Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (Accessed 19 Dec 2018).
20 Bax L, Yu LM, Ikeda N, *et al*. A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Med Res Methodol* 2007;7:40.
21 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539–58.
22 Kiran A, Crespillo AP, Rahimi K. Graphics and statistics for cardiology: data visualisation for meta-analysis. *Heart* 2017;103:19–23.
23 Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559–73.
24 Peters SA, Huxley RR, Woodward M. Diabetes as a risk factor for stroke in women compared with men: a systematic review and meta-analysis of 64 cohorts, including 775,385 individuals and 12,539 strokes. *Lancet* 2014;383:1973–80.
25 Kronish IM, Woodward M, Sergie Z, *et al*. Meta-analysis: impact of drug class on adherence to antihypertensives. *Circulation* 2011;123:1611–21.
26 Woodward M, Peters SA, Huxley RR. Diabetes and the female disadvantage. *Womens Health* 2015;11:833–9.