Human Reproduction, 2018
doi:10.1093/humrep/dey046

human reproduction

SUPPLEMENTARY DATA

Estimating the true proportion same-sex among dizygotic twins with misclassifications

We use Add Health for illustration because its estimated rate of misclassification (9%) is much greater than that for NCDS (1.7%). Thus, our estimate below for the potential effect of zygosity misclassification on our conclusion using Add Health is conservative and will be smaller for NCDS.

In Add Health Wave I data, 31% of the twins are classified as MZ and 69% are classified as DZ; 57% of DZ twins are SS and 43% are OS. Harris, Halpern, Smolen, and Haberstick (2006) note that the results of their DNA analysis at Wave III show that 9% of the twins were misclassified in their zygosity.

Out of 100 twins, let

s = number of true MZ twins correctly classified as MZ twins;
m = number of true MZ twins incorrectly classified as DZ SS twins;
S = number of true DZ SS twins correctly classified as DZ SS twins; and
M = number of true DZ SS twins incorrectly classified as MZ twins.

Of the 69 DZ twins, $69 \times 0.57 = 39.33$ are SS.

The number of twins classified as MZ:

$$s - m + M = 31 \quad (1)$$

The number of twins classified as DZ SS:

$$S - E + m = 39.33 \quad (2)$$

Total number of misclassifications:

$$m + M = 9 \quad (3)$$

The rates of misclassifications:

$$\frac{m}{s} = x\frac{M}{s} \quad (4)$$

where x = a scaler; $x = 1$ if MZ twins are equally likely to be misclassified as DZ SS twins as DZ SS twins are to be misclassified as MZ twins; $x > 1$ if MZ twins are more likely to be misclassified as DZ SS twins than DZ SS twins are to be misclassified as MZ twins; and $x < 1$ if DZ SS twins are more likely to be misclassified as MZ twins than MZ twins are to be misclassified as DZ SS twins.

We assume that DZ OS twins are never misclassified as MZ twins. Then there are $69 \times 0.43 = 29.67$ true DZ OS twins.

$$z = \frac{s}{s + 29.67} \quad (5)$$

where z = true proportion of SS twins among DZ twins.

We solve the simultaneous equations 1–4 for four unknowns (s, m, S, M) for various specified values of x, and then obtain z. Here are the results:

	x	z
MZ twins more likely to be misclassified as DZ SS twins	4.0	0.53621
	3.0	0.54385
	2.0	0.55716
	1.0	0.58281
DZ SS twins more likely to be misclassified as MZ twins	0.5	0.60229

Thus, under a wide range of conditions (specified by x), the proportion of SS twins among DZ twins (estimated by z) is always higher than 0.50000.