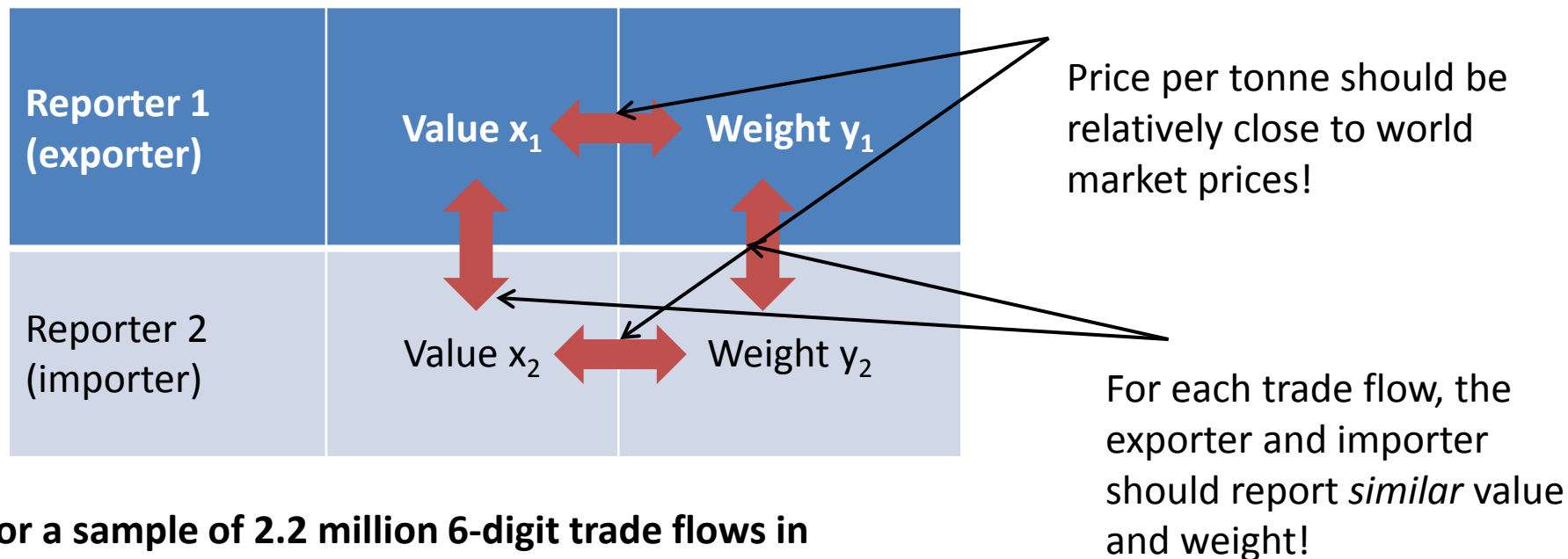


# The scope and scale of the challenge

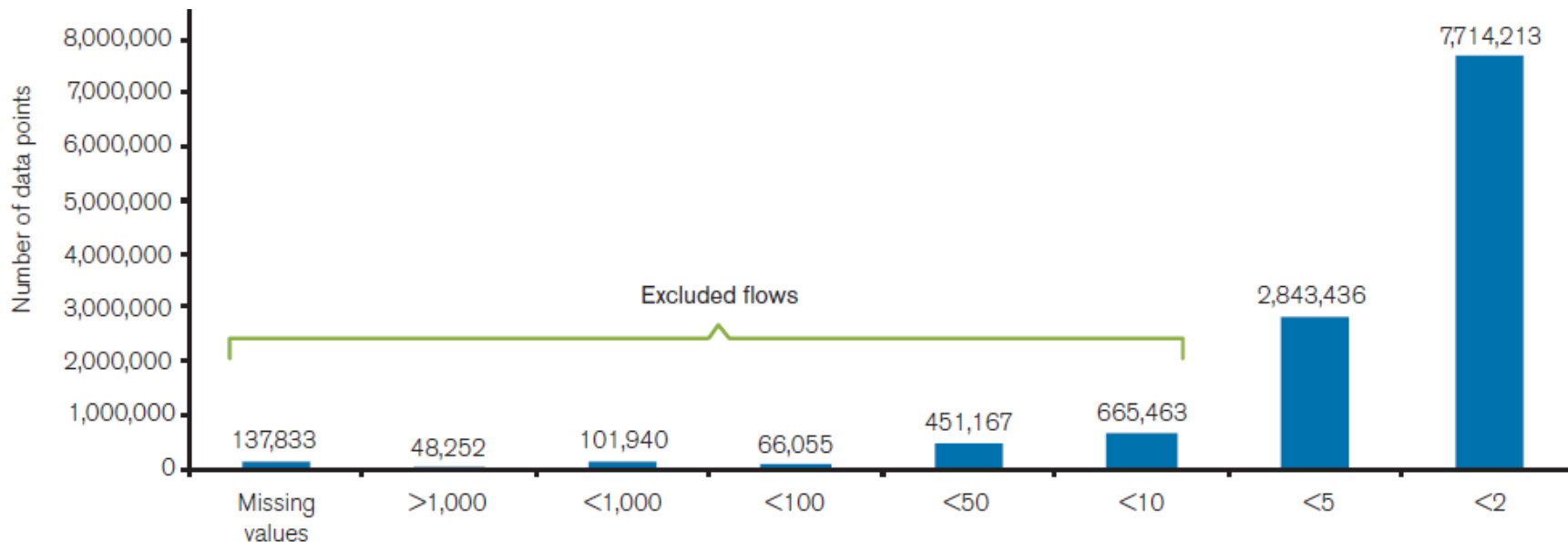


**For a sample of 2.2 million 6-digit trade flows in COMTRADE in 2011 and 2012**

- Only 37.3% had records on weight and values from both trade partners,
- of these, less than half (17.5%) are consistent with each other!
- Good news, these 17.5% contain >85% of the value and volume in these trade flows.

# BACI unfortunately is not the solution...

**BACI reconciles flows with a consistent method. But it relies on strong assumptions about reporter reliability, which are difficult to justify empirically, politically not palatable, and do not always improve the quality of the data.**

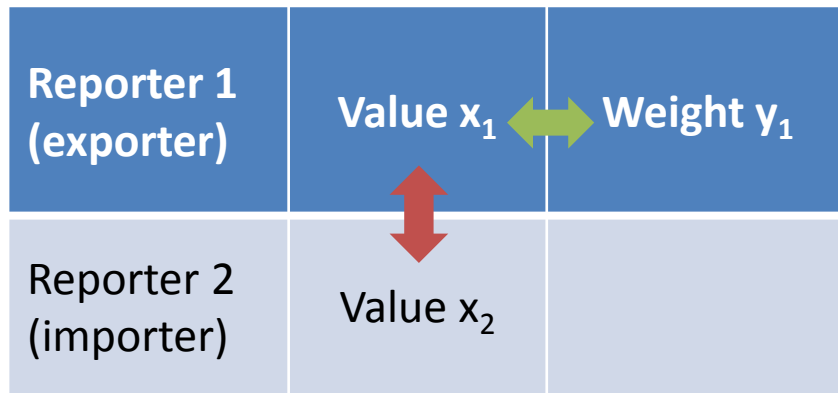


$$\text{If } \left( \frac{v_{t,hs,i_x,e_y}}{q_{t,hs,i_x,e_y}} \right) \geq \text{Median} \left( \frac{v_{t,hs,l,E}}{q_{t,hs,l,E}} \right), \text{ then } \text{Distance}_{t,hs,i_x,e_y} = \frac{\left( \frac{v_{t,hs,i_x,e_y}}{q_{t,hs,i_x,e_y}} \right)}{\text{Median} \left( \frac{v_{t,hs,l,E}}{q_{t,hs,l,E}} \right)}$$

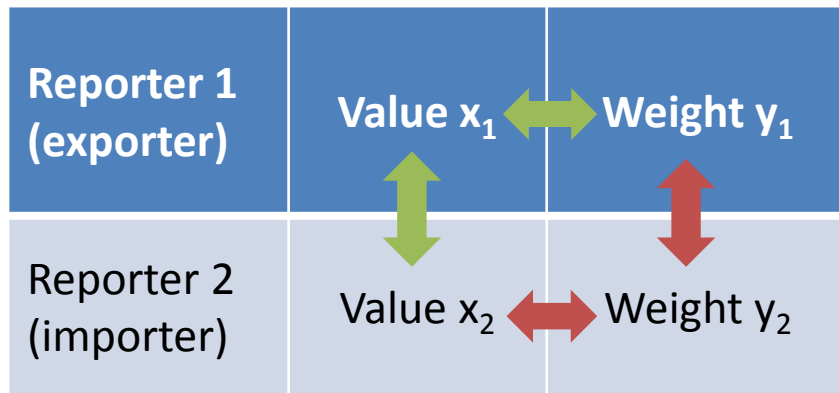
$$\text{If } \left( \frac{v_{t,hs,i_x,e_y}}{q_{t,hs,i_x,e_y}} \right) < \text{Median} \left( \frac{v_{t,hs,l,E}}{q_{t,hs,l,E}} \right), \text{ then } \text{Distance}_{t,hs,i_x,e_y} = \frac{\text{Median} \left( \frac{v_{t,hs,l,E}}{q_{t,hs,l,E}} \right)}{\left( \frac{v_{t,hs,i_x,e_y}}{q_{t,hs,i_x,e_y}} \right)}$$

Where  $l = i1, i2, i3, \dots$  represents the set of all importing countries;  $E = e1, e2, e3, \dots$  represents the set of all exporting countries;  $i_x$  and  $e_y$  are specific importing and exporting countries;  $t$  denotes the year; and  $hs$  the six-digit HS code of a particular resource product

# New intuitive and robust method for reconciling trade flows



=  
Value  $x_1$ ; Weight  $y_1$



=  
Average of Value  $x_1$  and  
Value  $x_2$ ; Weight  $y_1$

- Works at the level of the individual trade flow rather than more aggregate levels.
- Allows to salvage large share of data while avoiding strong assumptions about reporter reliability!

