Teoria delle Decisioni per SMID

paradosso di Simpson

La tabella coi numeretti è presa da Judea Pearl: *Simpson's Paradox: An Anatomy.*

Maschi	guarisce	non guarisce		tasso guarigione
medicina	18	12	30	60%
no medicina	7	3	10	70%
	25	15	40	
Femmine	guarisce	non guarisce		tasso guarigione
medicina	2	8	10	20%
no medicina	9	21	30	30%
	11	29	40	
M&F	guarisce	non guarisce		tasso guarigione
medicina	20	20	40	50%
no medicina	16	24	40	40%
	36	44	80	

Note:
1) non si faccia attenzione alla evidente ridotta numerosità di certi campioni. Possiamo moltiplicare per 1000, 1000000 i numeri dati e non cambia nulla del paradosso.
2) devo decidere se dare la medicina ad uno oppure no. Cosa faccio? Decido senza sapere se è maschio o femmina, o sapendolo? Ma, soprattutto, gliela do o no?

La soluzione alla domanda 2) è, per me, data dalla seguente tabella:

Maschi	guarisce	non guarisce		tasso guarigione
medicina	18	12	30	60%
no medicina	21	9	30	70%
	39	21	60	
Femmine	guarisce	non guarisce		tasso guarigione
medicina	6	24	30	20%
no medicina	9	21	30	30%
	15	45	60	
M&F	guarisce	non guarisce		tasso guarigione
medicina	24	36	60	40%
no medicina	30	30	60	50%
	36	44	120	

Ovvero, ho "riportato" i numeri assumendo che sia identico il numero di persone trattate o non trattate in ciascuno dei due campioni (sia M che F), e quindi anche nel campione complessivo (M e F)
Per cui io, ovviamente, non gli dò la medicina.
Non solo, ma in accordo col "sure-thing principle", visto che fa male sia a M che a F, è irrilevante sapere se ho davanti M o F.

E' corretta l'operazione fatta? Due considerazioni.
1) sarebbe buona norma cercare di fare sperimentazioni con "uguale numerosità"
2) la operazione di "moltiplicare" per 3 i casi che erano "minoritari" è giustificata se parto da campioni molto numerosi. Coi numeri piccoli (solo 10 maschi che non hanno preso la medicina e solo 10 femmine che l'hanno presa, farei una analisi un poco più attenta, parlando magari di intervalli di confidenza.