Allocating cases review

I think I still don’t fully understand how to correctly allocate infections in the absence of accounting for riskiness at work. As an example, let’s say we have two jobs, A and B, that are, in essence, equally risky. However, workers in job A are risk-takers who don’t wear a mask at work or at leisure, while workers in job B always wear a mask. Workers in job A also work 2 more hours. Let’s also assume that not wearing a mask triples your risk of a covid infection both in a work and leisure context. Following these rules, we could make the following hypothetical data:

Let’s say we know that workers in both job A spend 12 hours at home, 6 hours at work, and 6 hours at leisure:

Job A, 100 mask wearers

95 not infected

5 total infections:

* 1 leisure infection
* 2 work infections
* 2 home infection

Workers in job B spend 12 hours at home, 6 hours at work, and 6 hours at leisure

Job B, 100 mask wearers

96 not infected

4 total infections:

* 1 leisure infection
* 1 work infection
* 2 home infections

**Data**

The raw data corresponding to job A would have 100 total rows, 8 of which will have a 1 for the covid variable, indicating infection. As per our procedure, we split each row into three (a work respondent, a home respondent, and a leisure respondent). This data, combined with the procedure we outlined, would give us 6 of the following (total of 24 rows) for Job A (corresponding to the 6 infections):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Respondents | Weight | Job | No covid | Work covid | Home covid | Leisure covid |
| WorkResp | 1/3 | A | 0 | 0.75 | 0 | 0 |
| HomeResp | 1/3 | A | 0 | 0 | 1.5 | 0 |
| LeisureResp | 1/3 | A | 0 | 0 | 0 | 0.75 |

\*\*Here, I will flag that we decided to only apply a risky multiplier to the leisure infection.

And 92 of the following:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Respondents | Weight | Job | No covid | Work covid | Home covid | Leisure covid |
| WorkResp | 1/3 | A | 1 | 0 | 0 | 0 |
| HomeResp | 1/3 | A | 1 | 0 | 0 | 0 |
| LeisureResp | 1/3 | A | 1 | 0 | 0 | 0 |

For job B, we would have 4 of the following, corresponding to the infected mask-wearers:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Respondents | Weight | Job | No covid | Work covid | Home covid | Leisure covid |
| WorkResp | 1/3 | B | 0 | 0.75 | 0 | 0 |
| HomeResp | 1/3 | B | 0 | 0 | 1.5 | 0 |
| LeisureResp | 1/3 | B | 0 | 0 | 0 | 0.75 |

And 96 of the following:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Respondents | Weight | Job | No covid | Work covid | Home covid | Leisure covid |
| WorkResp | 1/3 | B | 1 | 0 | 0 | 0 |
| HomeResp | 1/3 | B | 1 | 0 | 0 | 0 |
| LeisureResp | 1/3 | B | 1 | 0 | 0 | 0 |

This would then yield the following expected values:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Job | HomeCovid | LeisureCovid | WorkCovid | NoCovid |
| A | 0.03 | 0.03 | 0.02 | 0.92 |
| B | 0.02 | 0.01 | 0.01 | 0.96 |

Which we would then fit to the known row and column totals:

* Population A: 100
* Population B: 100
* Home Covid: 4
* Leisure Covid: 3
* No Infection: 188
* Work Infection: 5

And obtain the following:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Job | HomeCovid | LeisureCovid | WorkCovid | NoCovid |
| A | 2.4 | 2.25 | 3.33 | 92 |
| B | 1.6 | 0.75 | 1.66 | 96 |

Because we don’t add risky multipliers to work, we are overestimating leisure and home infections and underestimating those happening at work. I think I’m doing something wrong, but I don’t quite understand in which part of the process I’m going astray. I would love your input!