Allocating cases review

I think my issue is not exactly with the “mandates enforced” assumption, but rather a more general confusion of how the procedure allows (or doesn’t allow) job risk to vary. Key point: the procedure I’m conducting is wrong because it doesn’t let one job be riskier than the other. Thus, I think I’m not following the steps you laid out correctly, and I wanted to see if you could tell me where I’m going wrong.

I’ll illustrate the issue with a simple example: Let’s say we have two jobs, A and B. Job A is twice as risky as job B (for reasons independent of mask-wearing, it could just be, for example, that one job requires closer contact with people). Additionally, workers in job A and B are all mask-wearers who spend 12 hours at home, 6 hours at work, and 6 hours at leisure.

Following these rules, we could make the following hypothetical data:

Job A, 100 mask wearers

95 not infected

5 total infections:

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

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, 5 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 25 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 |

And 95 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:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 |

Here, the issue becomes clear: infections for job A and job B have the same infection distribution across the three settings, even though work risk is different.

This would then yield the following expected values:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Job | HomeCovid | LeisureCovid | WorkCovid | NoCovid |
| A | 0.025 | 0.0125 | 0.0125 | 0.95 |
| 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: 2
* Work Infection: 3
* No Infection: 191

And obtain the following:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Job | HomeCovid | LeisureCovid | WorkCovid | NoCovid |
| A | 2.22 | 1.11 | 1.66 | 95 |
| B | 1.78 | 0.88 | 1.33 | 96 |

Even though job A is twice as risky, we are not accurately recovering this from the procedure. Instead, we are distributing this increased work risk across the other settings. This is obviously concerning to me because what we are most interested in is the fact that job A is twice as risky. Please let me know if this illustrates the issue!

As for how I thought of tackling this, my first instinct is not very elegant but serves as a preliminary idea. It follows the same process but adds a step at the beginning to calculate the job multiplier. To go back to the same example:

Job A, 100 mask wearers

95 not infected

5 total infections:

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

Job B, 100 mask wearers

96 not infected

4 total infections:

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

Of course, we don’t see the distribution of infections within jobs. Instead, this is all the information we have:

* Workers in both jobs spend 12 hours at leisure, 6 hours at work, and 6 hours at home.
* Workers in both jobs wear a mask during leisure.
* Total infections in job A = 5
* Total infections in job B = 4

This information is enough to obtain a work risk multiplier. Based on infection distribution, we know the ratio of infections is 1.25 (5/4). We thus just need to determine what job risk multiplier gets us this overall risk ratio. Because home and leisure should be equally risky for people in these jobs, we know that 0.75 of their day is equally risky, so we need a 2x multiplier for the remaining 0.25 of the day:

1(0.5) + 1(0.25) + x(0.25) = 1.25

x = 2

We use this multiplier when creating our three subrespondents. For job A, we’d have 5 of the following:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Respondents | Weight | Job | No covid | Work covid | Home covid | Leisure covid |
| WorkResp | 1/3 | A | 0 | 1.2 | 0 | 0 |
| HomeResp | 1/3 | A | 0 | 0 | 1.2 | 0 |
| LeisureResp | 1/3 | A | 0 | 0 | 0 | 0.6 |

The rest are just as described above. Now, we obtain the following expected values:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Job | HomeCovid | LeisureCovid | WorkCovid | NoCovid |
| A | 0.02 | 0.01 | 0.02 | 0.95 |
| B | 0.02 | 0.01 | 0.01 | 0.96 |

And correctly recover the number of cases in each setting x job:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Job | HomeCovid | LeisureCovid | WorkCovid | NoCovid |
| A | 2 | 1 | 2 | 95 |
| B | 2 | 1 | 1 | 96 |

Now, confirming the procedure works with a more complicated case. Above, we assumed that, at a baseline, all settings are equally risky­­–i.e., if a person wears a mask, then an hour at leisure is equally risky than an hour at work. This assumption does not need to hold, the only assumption we need to make is that if a person in job A wears a mask and a person in job B wears a mask, 1 hour at leisure is equally risky for them.

A new example:

* the distribution of hours in the day is the same (12 hours at home, 6 hours at work, 6 hours at leisure)
* workers in job B don’t wear a mask and their leisure hours are twice as risky
* job A is twice as risky as job B (though we wouldn’t know this)
* rates of infection at each setting are not equal (1 hour at home is not equally risky as 1 hour during leisure)

Job A, 100 mask wearers

60 not infected

40 total infections:

* 10 leisure infection
* 26 work infections
* 4 home infections

Job B, 100 non-mask wearers

63 not infected

37 total infections:

* 20 leisure infection
* 13 work infection
* 4 home infections

Again, we don’t see the distribution of infections within jobs. Instead, this is all the information we have:

* Workers in both jobs spend 12 hours at leisure, 6 hours at work, and 6 hours at home.
* Workers in job B don’t wear a mask during leisure.
* Total infections in job A = 40
* Total infections in job B = 37

We start with calculating the work multiplier:

1(0.5) + 0.5(0.25) + x(0.25) = 40/37

0.25x = 0.456

x = 1.82

18x = 30

x = 1.66

Using this to create the “infection profile” for those in job A. We have 40 of the following:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Respondents | Weight | Job | No covid | Work covid | Home covid | Leisure covid |
| WorkResp | 1/3 | A | 0 | 1.134 | 0 | 0 |
| HomeResp | 1/3 | A | 0 | 0 | 1.244 | 0 |
| LeisureResp | 1/3 | A | 0 | 0 | 0 | 0.622 |

For those in job B, we have 37 of the following:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Respondents | Weight | Job | No covid | Work covid | Home covid | Leisure covid |
| WorkResp | 1/3 | A | 0 | 0.6 | 0 | 0 |
| HomeResp | 1/3 | A | 0 | 0 | 1.2 | 0 |
| LeisureResp | 1/3 | A | 0 | 0 | 0 | 1.2 |

Our expected values:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Job | HomeCovid | LeisureCovid | WorkCovid | NoCovid |
| A | 0.02 | 0.01 | 0.02 | 0.95 |
| B | 0.02 | 0.01 | 0.01 | 0.96 |