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. In short: I believe 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 5 of the following (total of 15 rows) for Job A (corresponding to the 5 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 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!

The way I’m thinking about it is that, if we are committed to making the assumption that 1 hour of leisure for workers in job A will be equal to 1 hour of leisure for workers in job B, assuming they have the same mask behavior and a similar assumption to hours at home, then what we’d need is closer to a system of equations than an IPF procedure. To give another example:

Job A

* 12 home infections
* 32 leisure infections
* 25 work infections

Job B

* 9 home infections
* 8 leisure infections
* 30 work infections

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

* Total workers job A = 400
* Total workers job B = 200
* Workers in job A spend 8 hours at home, 6 hours at leisure, and 10 hours at work
* Workers in job B spend 12 hours at home, 6 hours at leisure, and 6 hours at work
* Workers in job A don’t wear a mask
* Total infections in each setting
* We also know, from the Pulse survey, that job A has an infection rate that is 1.08 times higher than that in job B

Putting this information into a table where we multiply the unknown covid rates for each setting by the number of hours spent there, add a risk multiplier for leisure (assuming not wearing a mask makes each hour twice as risky although we know this isn’t the right number), and multiply rates by 2 in job A because the population is twice as large:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Job | HomeCovid | LeisureCovid | NoCovid | Total Population |
| A | 2\*8H | 6L \* 2 \* 2 | 2\*1.08X | 400 |
| B | 12H | 6L | X | 200 |
|  | Total = 21 | Total = 40 | Total = 484 |  |

Then, it’s just a matter of solving for each individual rate to get at the total number of work infections:

H = 21/28

L = 40/30

X = 484/3.16

For job A:

Work infections + 16(21/28) + 24(40/30) + 2.16(484/3.16) = 400

Work infections = 25

For job B:

Work infections + 12(21/28) + 6(40/30) + 1(484/3.16) = 200

Work infections = 30

I’m not sure if this makes sense and whether I’m just misunderstanding the IPF procedure! The bottom line for me is that, since we are thinking of work infections as “the balance” of the other two types of infections, then we don’t want to set any initial distribution for work risk rates before conducting the IPF procedure. Instead, we want to leave work infections as kind of a “free variable” that we set based on information from the other two settings.