# Crosswalks Between (D)ISCO88 and (D)ISCO08 Occupational Codes

Anders Humlum*

*University of Chicago*

September 4, 2021

## 1 Introduction

The International Standard Classification of Occupations (ISCO) updated in 2010 their four-digit classification from ISCO88 to ISCO08. Statistics Denmark (DST) simultaneously updated their six-digit disaggregation from DISCO88 to DISCO08.

In this note, we tackle the problem that there exist no conversion keys between (D)ISCO88 and (D)ISCO08. The International Labour Organization (ILO) has published an `m:m` correspondence table from ISCO88 to ISCO08. DST has published a partial `1:m` conversion key for DISCO88 codes with clear correspondence to DISCO08 codes.[1] We combine these correspondence tables with microdata on observed code reassignments within job spells to infer `m:1` crosswalks between DISCO08 and DISCO88 codes.

In Section 3, we analyze the constructed crosswalks. We show that the crosswalks dramatically improves the stability of occupational transition rates around the change in nomenclature.

The crosswalks are available for download here.

*E-mail: humlum@uchicago.edu. I thank Johannes L. R. Hansen for superb research assistance.

[1] As discussed in this note, there exists no complete DISCO88 to DISCO08 conversion key.

# 2 Constructing the Crosswalks

To assist the construction of crosswalks between DISCO88 and DISCO08, we study workers who stay at the same establishment across the data break in 2010.[2] Following these workers, we can assess empirically how DISCO88 codes are reassigned to DISCO08 codes within job spells.[3]

In principle, one could just link each DISCO88 code to the DISCO08 code with most workers, registered initially under the DISCO88 code, who stay in their jobs and get reassigned to the DISCO08. There are, however, multiple challenges to this automated approach. We discuss these challenges in Section 2.1 before proposing our linking method in Section 2.2.

## 2.1 Challenges

### 2.1.1 Occupational Titles Added or Discontinued

In the constructing `m:1` crosswalks, we face the challenge that, when the nomenclature was updated in 2010, some codes were split into multiple, more detailed codes, while other codes were aggregated to one, more general code. At the same time, some new occupations were added, and some outdated ones were discontinued. This makes it impossible to construct true `m:1` crosswalks between DISCO88 and DISCO08. Instead, we construct `m:1` crosswalks where the best match is found for each code. This creates some ambiguity in selecting the correct code, as discussed below.

An illustrative example is that of military occupations. In the DISCO88 nomenclature, there is only one category: *Military work*. In the DISCO08 nomenclature, on the other hand, there are three categories: *Military work on officer level*, *Military work on under officer level*, and *Military work, other ranks*. In this case, it is not clear which of the three DISCO08 occupational categories the DISCO88 code should be linked to.

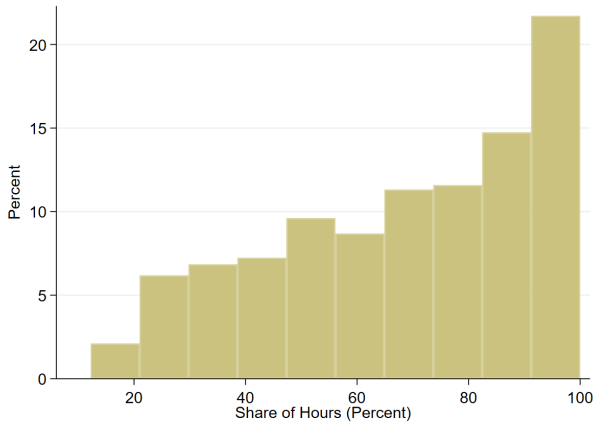
Figure 1 shows this ambiguity empirically. For each DISCO88 code, Panel (a) plots the highest share of workers reassigned to a DISCO08 code. For many DISCO88 codes, this "top DISCO08 candidate" share is rather low, making it unclear whether it indeed is the correct link.

[2] We use the Employer-Employee Register (BFL) which contain high-quality and comprehensive information on job spells, occupational codes, and hours of workers in Denmark.

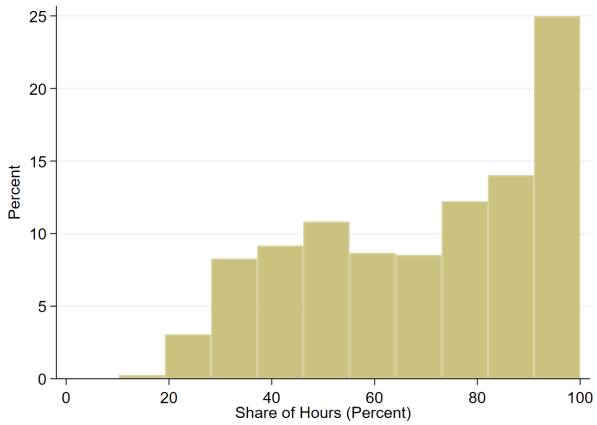
[3] Our strategy is inspired by Heinesen et al. (2018); see their Appendix A2.1.

**Figure 1:** Share of Hours For Top Candidate

**(a)** DISCO08-to-DISCO88




**(b)** DISCO88-to-DISCO08




### 2.1.2 Spurious Continuations of Numerical Occupation Codes

A second challenge is that not all companies updated the numerical occupation codes they assigned to employees between 2009 and 2010, even though the nomenclature changed. These code copies create some bizarre, spurious links, where the numerical code is the same in both years, but the occupations assigned to the code in different years are very different. One example is "Butcher and treatment of fish" (DISCO88 `741100`) being reassigned to "Electrician" (DISCO08 `741100`) within a job spell. These copies are particularly troublesome in cases where the occupations assigned to a code in different nomenclatures differ in size, such that the spurious links dominate the correct matches.

Taking these pitfalls to blindly relying on observed reassignments of occupational codes, we construct our crosswalks using the method described in the next section.

## 2.2 Linking Method

We construct four `m:1` crosswalks (two DISCO88-DISCO08 and two ISCO88-ISCO08). The ISCO crosswalks are constructed using the same methodology as DISCO, just shortening the codes to four digits before applying the procedure.

To assist the linking of occupational codes across nomenclature, we study workers who stayed in their job spell from 2009 to 2010. For example, to link DISCO88 to DISCO08, we calculate, for each DISCO88 in 2009, the share of workers that are reassigned to each

DISCO08 code (weighted by hours of employment).[4] This gives us a list of potential candidates DISCO08 codes for each DISCO88 code, ranked by their actual usage as a replacement for the DISCO88 code. To narrow down this list of candidate matches, we use the following four criteria:

1. The code has the highest share of hours worked of any of the candidates and is included in the ILO/DST conversion tables.

2. The code's share of hours worked is less than 20 percentage points lower than the link with the highest share of hours worked, and it is in the ILO/DST tables.

3. The code's share of hours worked is more than 20 percentage points higher than the second most used code, and the link is based on more than 50 observations.

4. The link is the only candidate found and the link is based on more than 50 observations.

To address the issue of companies not updating codes between 2009 and 2010 (Section 2.1.2) and other abnormalities with sparsely used codes, we manually check all automatically generated candidates and remove spurious links.

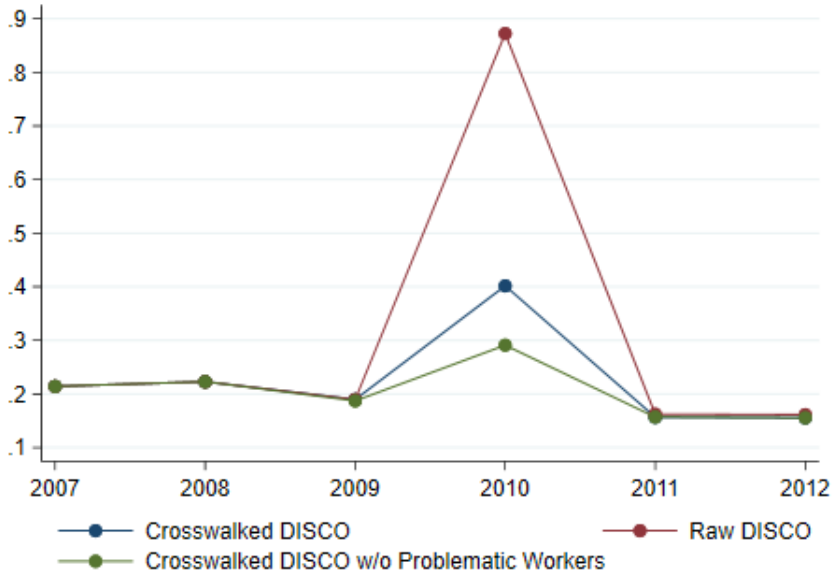
Finally, we choose manually among the codes that satisfy the criteria above, trading off the code's share of employment against the comparability of occupational titles. For a few sparsely used codes, there exists no candidate satisfying the criteria above. In these cases, we pick the crosswalk codes manually from the entire catalogue of codes.

# 3 Analyzing the Crosswalks

In this section, we evaluate the validity of our constructed crosswalks. Figure 2 plots the share of workers who change their six-digit occupation code each year. As expected, an abnormally large portion of the workforce changed their raw DISCO code when the nomenclature was updated from 2009 to 2010. After converting the codes after 2010 to the DISCO88 nomenclature, the jump is less pronounced. However, as we discuss below, the discontinuity will not disappear entirely even with an ideal crosswalk.

[4] We link each DISCO08 to a DISCO88 code using the same methodology, now just based on the reassignment share of hours worked.

**Figure 2:** Share of Workers Changing 6-Digit Occupation Code




The two challenges discussed in Section 2.1 contribute to the observed jump in occupational switching around the change of nomenclature. First, if two DISCO88 codes are combined into a single DISCO08 code, our `m:1` DISCO08-to-DISCO88 crosswalk will link all of the workers to a single ISCO88 code after 2010. To eliminate this source of discontinuity, the green line in Figure 2 excludes workers who were initially employed in discontinued DISCO88 codes.[5,6] Second, if numerical codes were automatically copied between 2009 and 2010, even though the nomenclature changed, the crosswalk would spuriously infer that workers changed occupations. To eliminate this source of discontinuity, the green line in Figure 2 excludes DISC88-DISCO08 transitions where the numerical code is identical, but the associated occupations differ.[7]

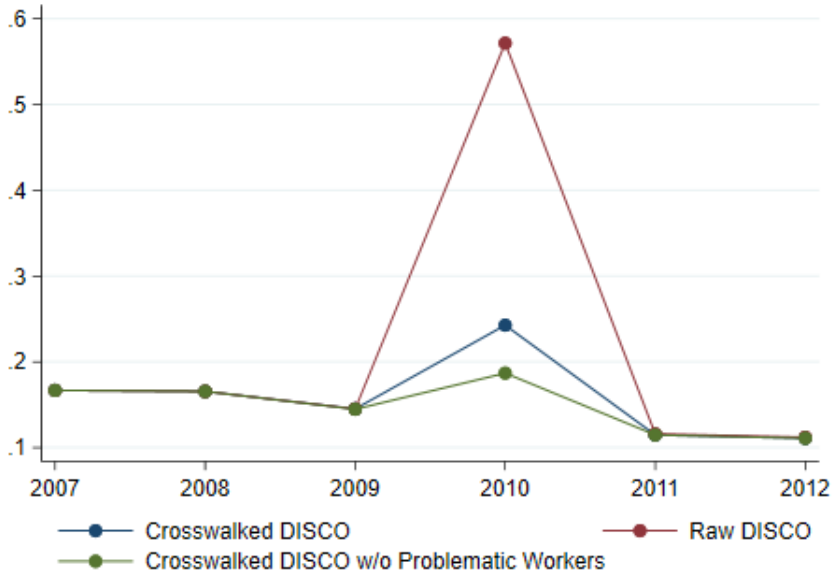
Figure 3 shows the share of workers who change their two-digit occupation code each year. As the figure shows, most of the nomenclature revisions from DISCO88 to DISCO08 happened within two-digit categories.

[5] A DISCO88 code is discontinued if it does not appear in our `m:1` DISCO08-to-DISCO88 crosswalk.

[6] With panel data on job spells, the spurious occupational switching due to discontinued codes can be resolved as follows: (*i*) For all discontinued DISCO88 codes `X`, find their link `Z` in the DISCO88-to-DISCO08 crosswalk, (*ii*) if a worker is employed in `X` before 2010, continue to assign him to code `X` after 2010 as long as he stays in the job spell and now has DISCO08 code `Z`.

[7] With panel data on job spells, this problem can be resolved by keeping a worker's initial DISCO88 code if he stays in the job spell, and his numerical DISCO88 and DISCO08 codes are identical despite these relating to different occupations.

**Figure 3:** Share of Workers Changing 2-Digit Occupation Code




Finally, we conjecture that the jump in occupational switching in 2010 partly could be due to the change in nomenclature prompting employers to reevaluate their assignment of occupational codes. Groes et al. (2015) show that employers rarely update their occupation codes within job spells.

# References

Groes, F., P. Kircher, and I. Manovskii (2015). The U-Shapes of Occupational Mobility. *The Review of Economic Studies 82*(2), 659–692.

Heinesen, E., S. Imai, and S. Maruyama (2018). Employment, Job Skills and Occupational Mobility of Cancer Survivors. *Journal of Health Economics 58*, 151–175.