

UCD SNOMED CT Project *

SNOMED CT Concept Assignment Frequency

Version 2.0, September 8, 2020

Tom Balmat, Duke University Research Computing

Following are various plots of the distribution of SNOMED CT concept assignment frequency as observed in our data. Section 1 contains distributions of the number of participant assignments per concept. Section 2 contains distributions of the number of concepts assigned by various joint categories of participant covariates.

1 Distribution of assignments per concept

Figure 1 shows the overall distribution of participant assignment frequency per concept. The x-axis extends to 224, the maximum number of assignments for any concept.¹ Figure 2 limits the x-axis to thirty or fewer assignments, to zoom in on concepts assigned once, twice, three times, etc. An interesting observation is that most (2,932 of 5,070) concepts are assigned once only (to a single participant).

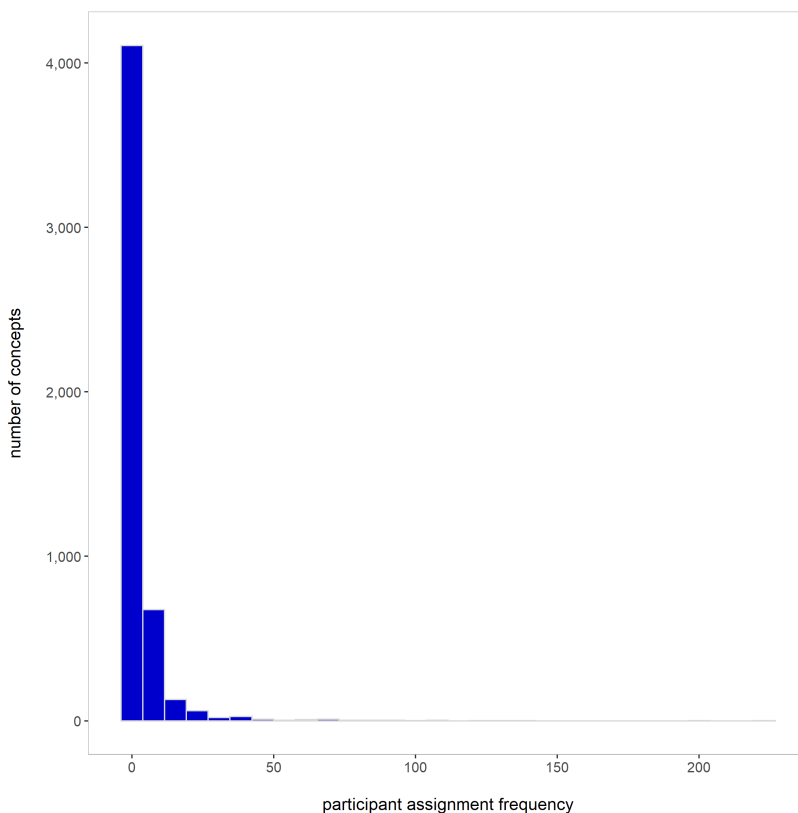


Figure 1: Distribution of number of assignments per concept

*Duke University, University of Michigan, University of Nebraska, Children's National Health System, Baylor University

¹The concept assigned to 224 participants happens to be "Ornithine carbamoyltransferase deficiency (disorder)"

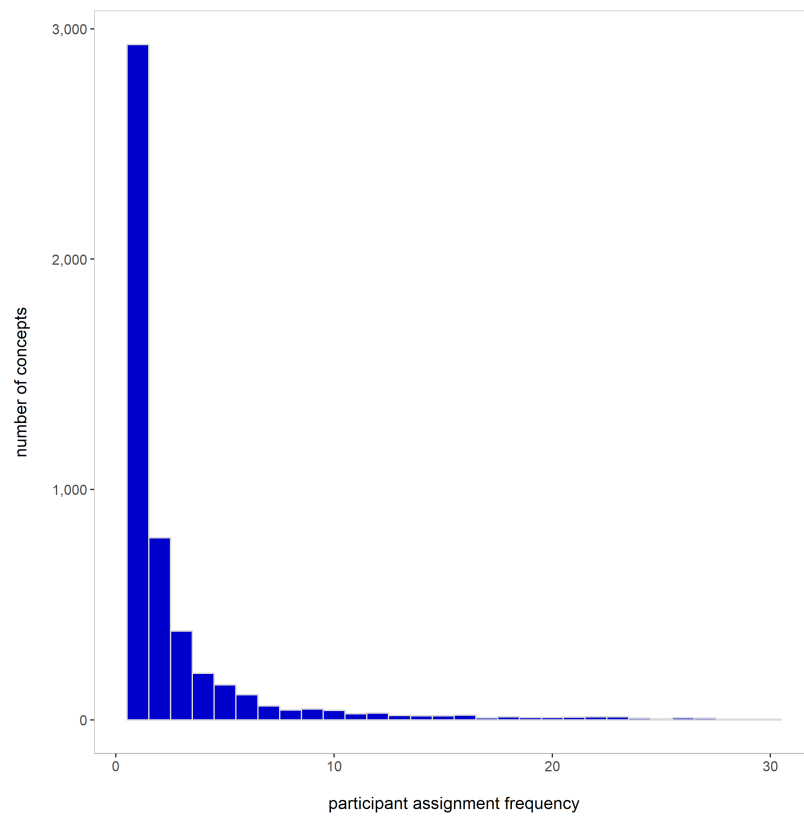


Figure 2: Distribution of number of assignments per concept, assignments of thirty or less

2 Distribution of concepts assigned per participant

Figure 3 shows the distribution of the number of concepts assigned per participant. Twenty-six participants are assigned a single concept, while one participant is assigned 145 concepts (the maximum assigned). The median number of concepts assigned per participant is seventeen. Figures 4 through 13 show the distribution of participants by various joint categories of covariates. Observations:

- Figure 4 indicates a difference in distribution of females by proximal-distal category, with an apparent significant shift (lower) in number of concepts assigned per proximal female. This should be considered when assessing associations involving proximal and sex covariates.
- Figure 5 indicates confounding of proximal-distal category and UCD diagnosis, since diagnoses ALD, ARG, ASD, CITR, and HHH appear strictly with category distal, while diagnoses CPS1, NAGS, and OTC appear strictly with category proximal. Perhaps there are biological or medical reasons for confounding, but it should be considered when assessing the relationship of UCD diagnosis or proximal-distal category to other covariates since, with our data, there does not appear to be a way of separating the effects of these variables.
- Although figure 6 shows a visually similar distribution by level of HA, it also indicates a difference in distribution by proximal-distal category, independent of HA. In addition to an apparent slight increase in proximal participants, concept diversity appears decreased for these participants (both HA levels), since mass is shifted left, to lesser concepts assigned, on corresponding x-axes.
- Figure 7 indicates under-representation of participants of age 11 to 100 days, especially for category proximate. This should be considered when assessing associations of age with other covariates.
- Figure 8 indicates over-representation of females for diagnosis OTC.
- Figure 9 indicates reduced diversity in concepts for non-HA, particularly for females.
- Figure 10 indicates an over-representation of females in age categories greater than 100 days.
- Figure 11 indicates that most participants have UCD diagnoses of OTC, ALD, or ASD. Participants appear uniformly distributed (for these diagnoses) between HA and non-HA status, although non-HA exhibits less diversity in concept assignment.
- Figure 12 indicates under-representation of ALD and ASD, or over-representation of OTC, participants in age categories greater than 100 days..
- Figure 13 indicates over-representation of HA participants of age zero to eleven days and over-representation of non-HA participants of age greater than 10,000 days.

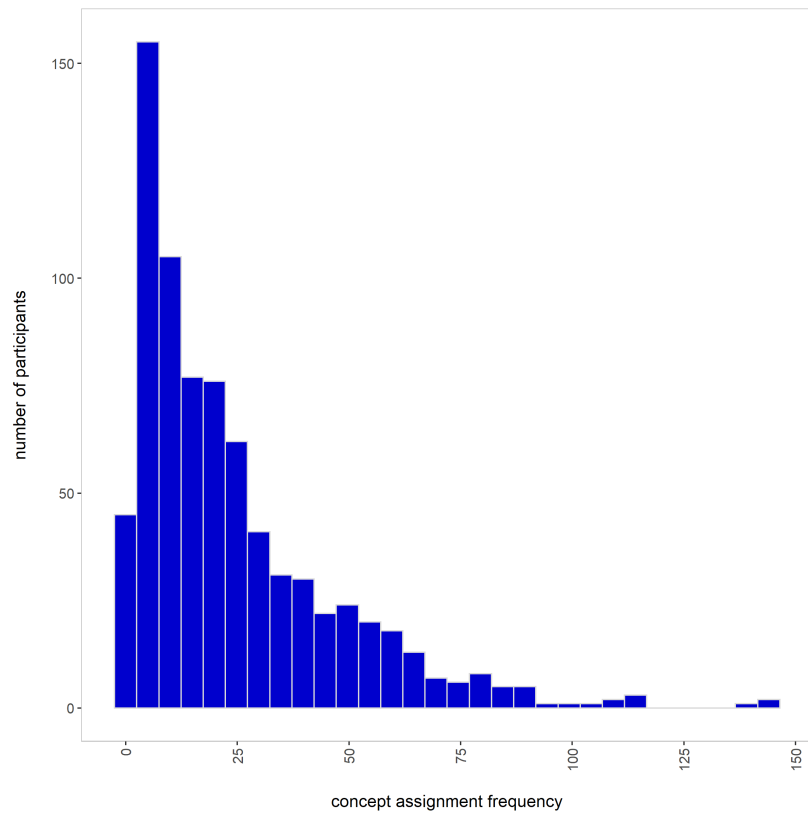


Figure 3: Distribution of number of concepts assigned per participant

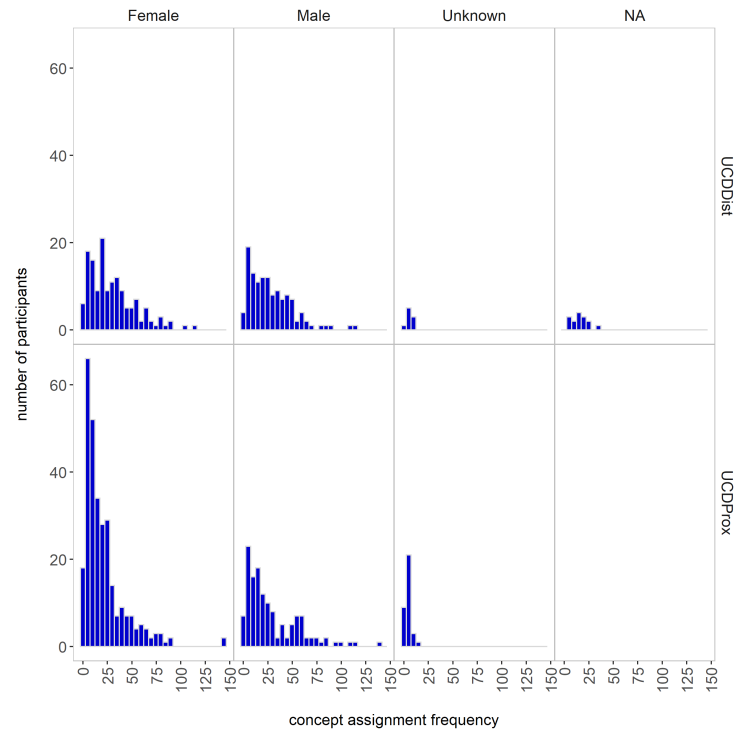


Figure 4: Distribution of number of concepts assigned per participant, by proximal-distal category and sex

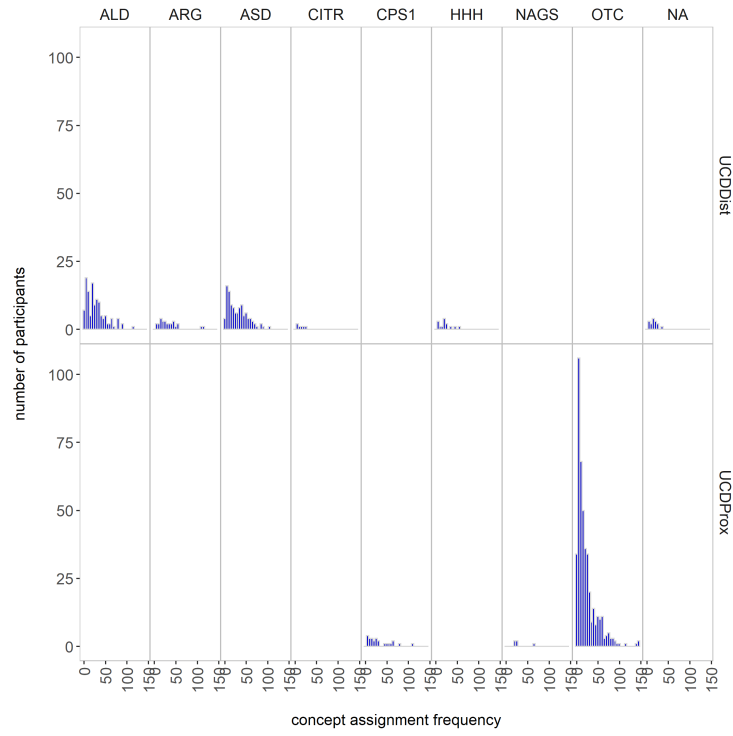


Figure 5: Distribution of number of concepts assigned per participant, by proximal-distal category and UCD diagnosis

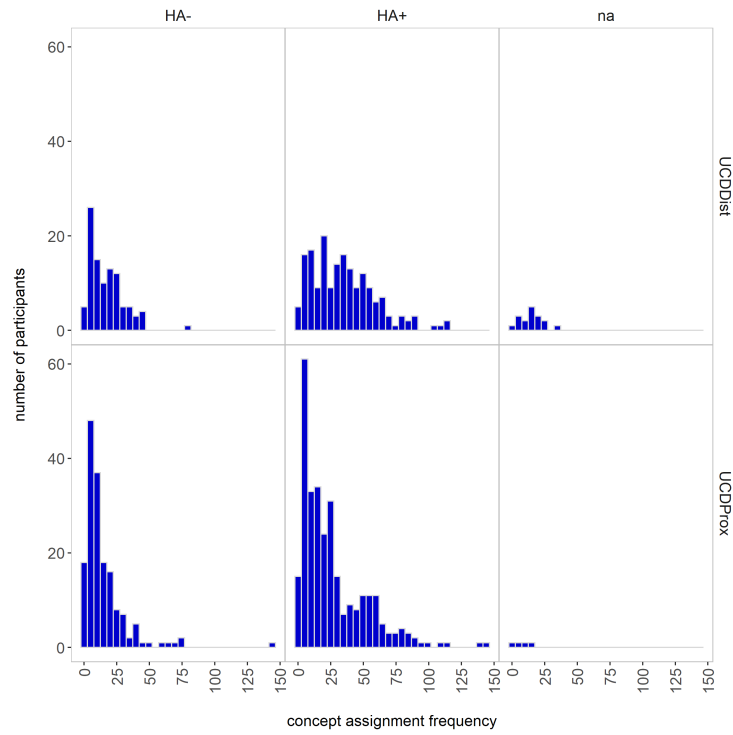


Figure 6: Distribution of number of concepts assigned per participant, by proximal-distal category and occurrence of HA events

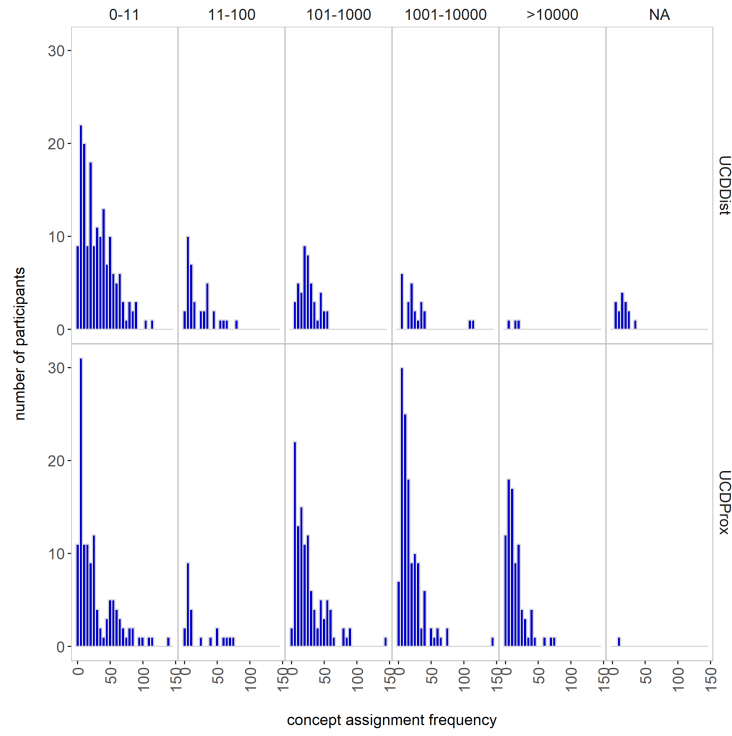


Figure 7: Distribution of number of concepts assigned per participant, by proximal-distal category and age (in days)

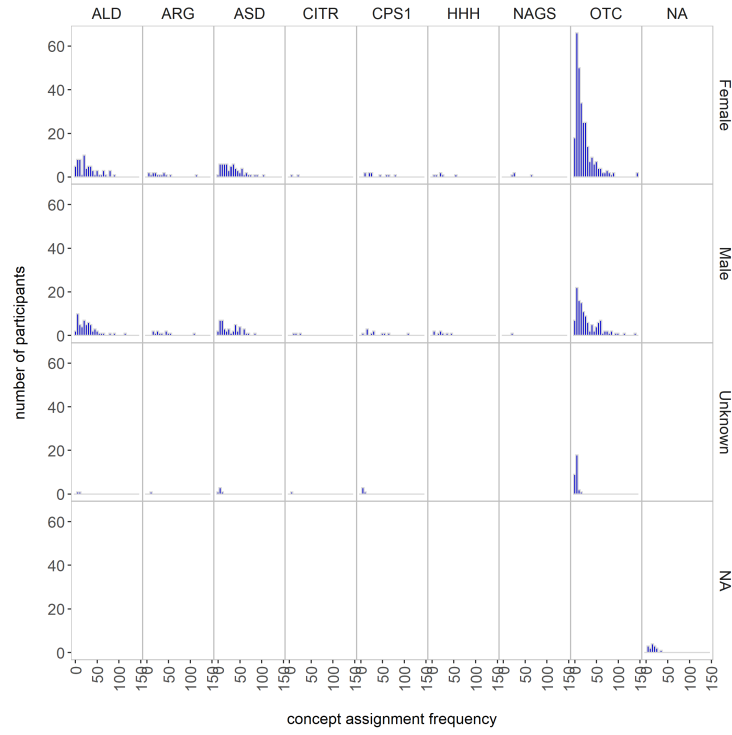


Figure 8: Distribution of number of concepts assigned per participant, by sex and UCD diagnosis

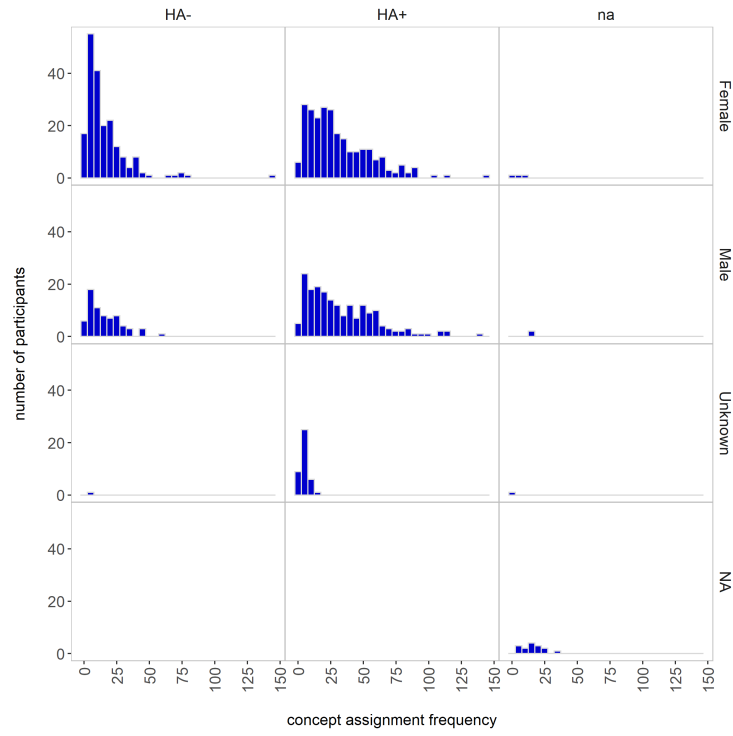


Figure 9: Distribution of number of concepts assigned per participant, by sex and HA

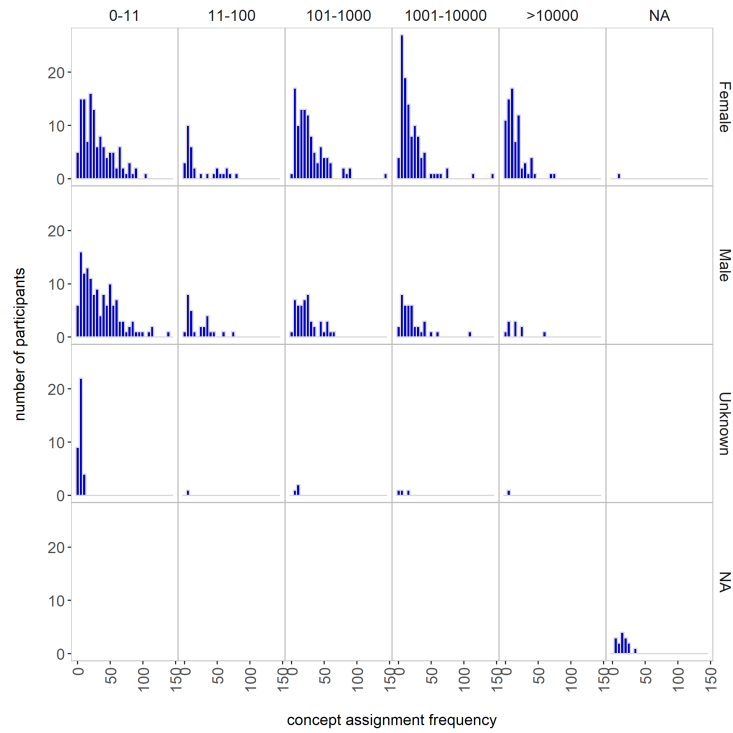


Figure 10: Distribution of number of concepts assigned per participant, by sex and Age (in days)

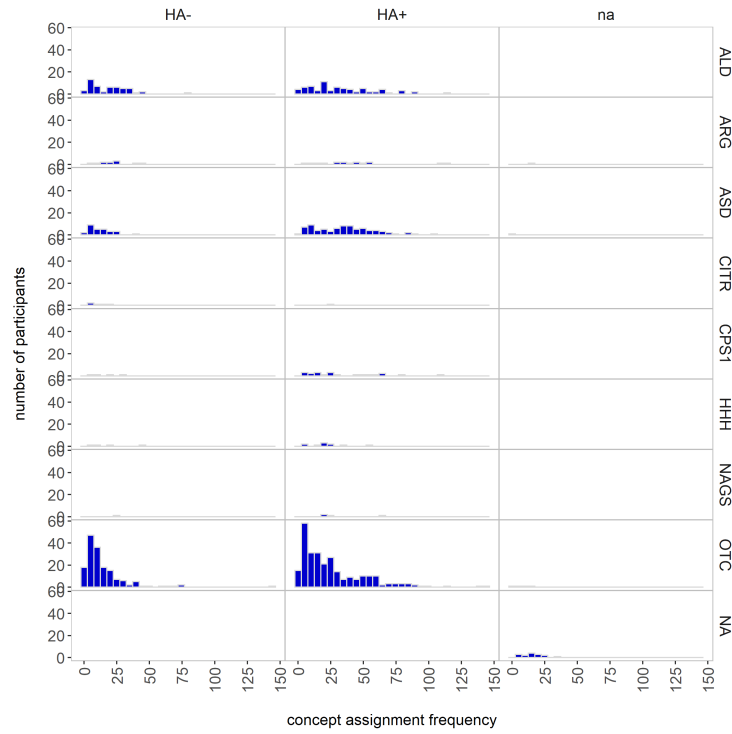


Figure 11: Distribution of number of concepts assigned per participant, by UCD diagnosis and HA

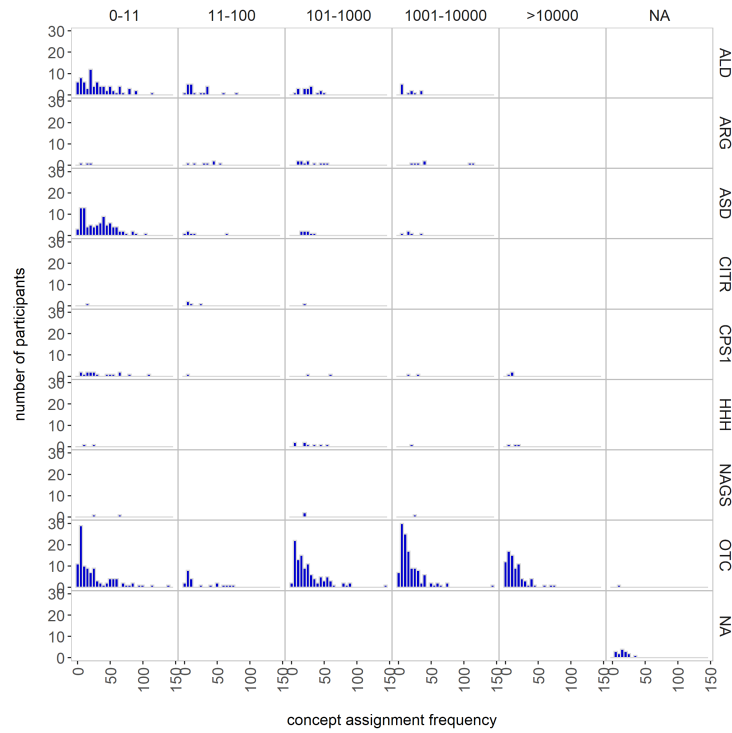


Figure 12: Distribution of number of concepts assigned per participant, by UCD diagnosis and Age

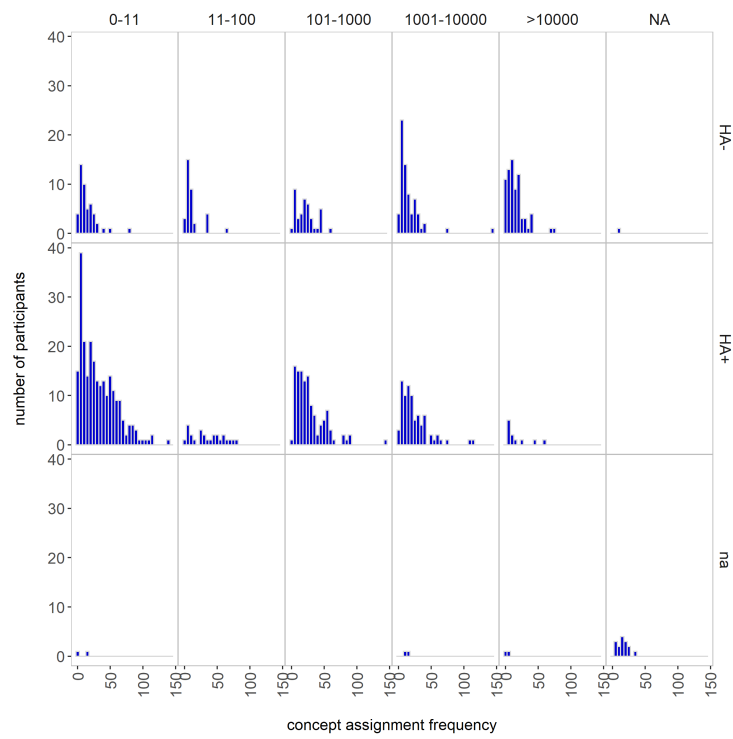


Figure 13: Distribution of number of concepts assigned per participant, by HA and Age