UK Health Research Analysis 2014: Supplementary Information

Supplementary Information #02: Automated Coding of Semantic HRCS Health Categories

The ÜberResearch team, using their tool Dimensions, have integrated a first iteration of HRCS HC codes using semantic categories. Dimensions has a curated grants database from over 90 funders worldwide, and therefore they can thread their semantic classifications through data sets that could not be examined by the manual approach. These insight help expand our understanding as they place the UK activities into a global perspective.

ÜberResearch will use the data in our manually coded grants to improve the accuracy rate of these HRCS HC codes and will be implementing the RAC codes during 2015.

Giles Radford

Head of Professional Services, ÜberResearch Ltd.

Section 1 – Activity in other RCUK councils

One of the issues facing HRAF reporting using HRCS codes is that some funders are not using the system, meaning we see an incomplete picture of UK activity. This is particularly true of the RCUK, where we know that some EPSRC, AHRC, ERRC and NERC grants have research activity that could be coded to HRCS, but given the bulk of their activity falls outside of biomedical research these grants are not been coded manually and total funding is therefore under-represented.

Figure 1a below shows total RCUK HRCS HC activity, coded by Dimensions (using their first attempt at HRCS coding and allowing for double counting where a grant falls into more than 1 category). Figure 1b then shows the same data with MRC and BBSRC removed.

Obviously the MRC and BBSRC are the largest funders, but, for grants starting 1/1/2010 – to 31/12/2014, there's about £400m of grants from these other councils coded to HRCS, ignoring Generic Health Research. There may be some double counting, so this estimate may be high, and there may be some questionable activity (questionable as to whether it is either research or affecting human health) but it still represents considerable activity worth further investigation.

One very interesting aspect of threading HRCS codes into areas outside of biomedical research is that not all that is returned is necessarily appropriate, but neither are they entirely inappropriate. Economic or Social Science grants clearly contribute to combatting issues, such as Mental Health, but they may not be what we might consider 'research'. Equally, NERC and EPSRC may fund activity that is more about the health of the world's flora and fauna, but is that really what HRCS was meant to report? So the HRAF group may have to consider how to handle these additional grants.
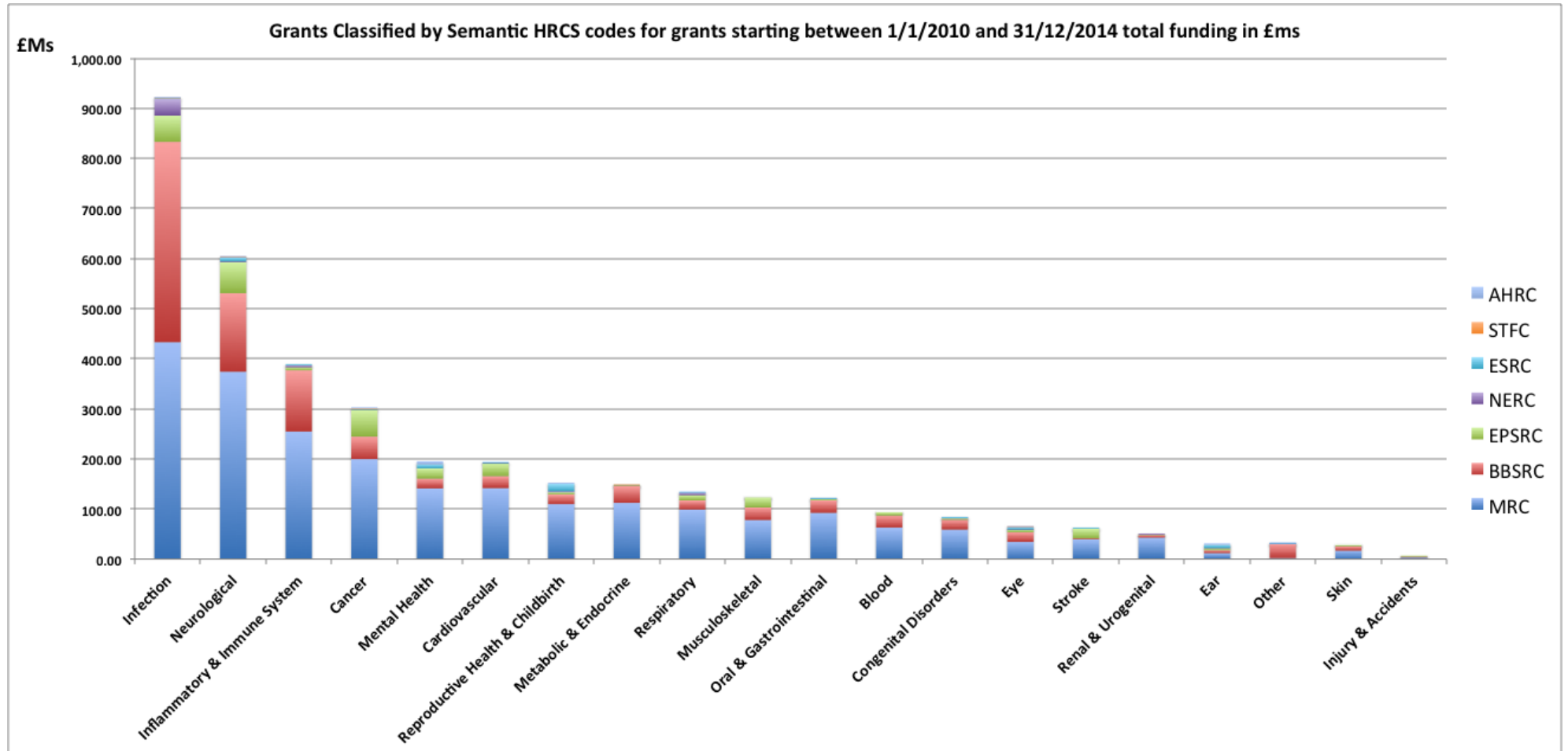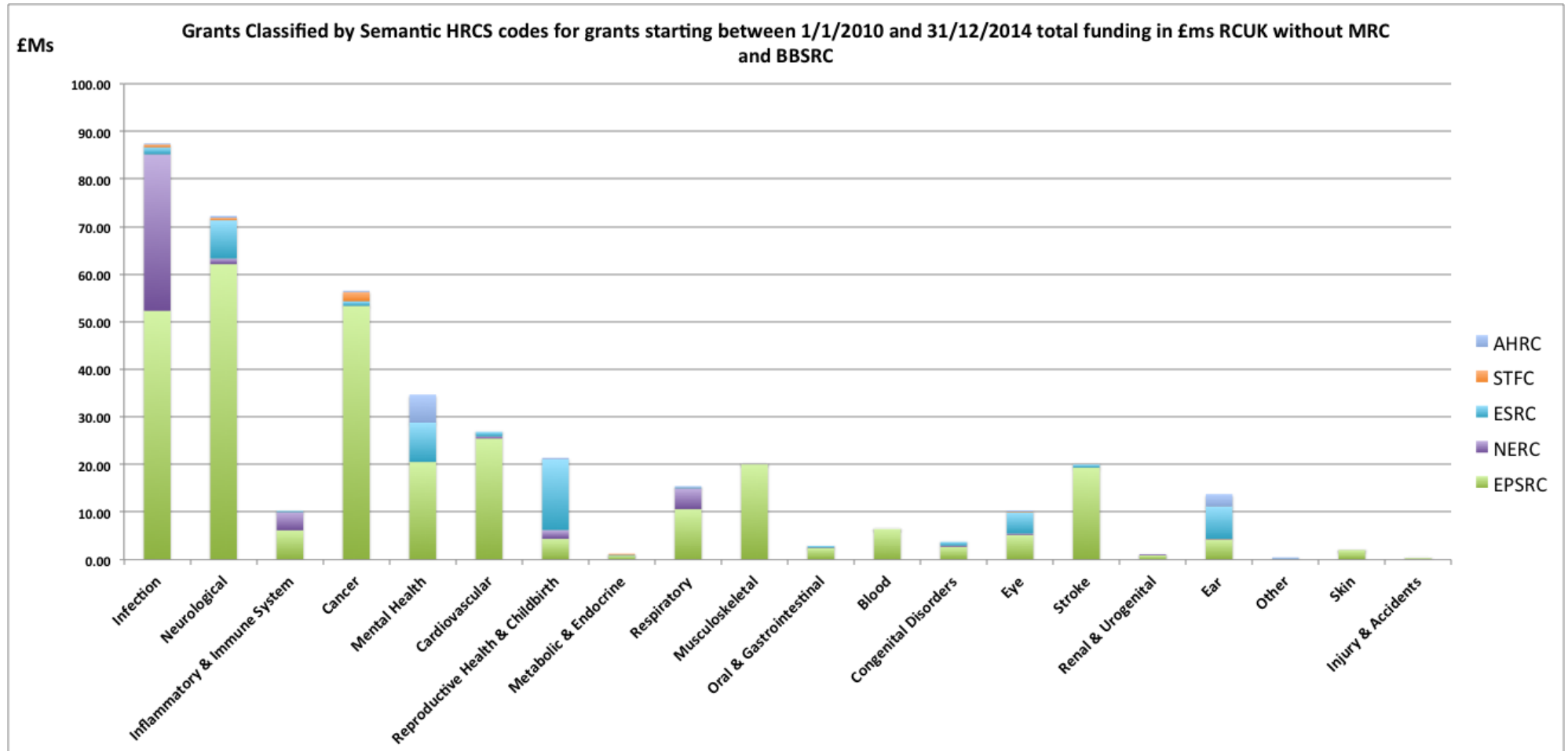
Figure 1a

Grants Classified by Semantic HRCS codes for grants starting between 1/1/2010 and 31/12/2014 total funding in £ms

Figure 1b

Grants Classified by Semantic HRCS codes for grants starting between 1/1/2010 and 31/12/2014 total funding in £ms RCUK without MRC and BBSRC

Section 2 – Government funding activity in the UK compared to activity in the USA

Understanding the total activity compared to the largest player in biomedical funding, the NIH, can highlight some interesting areas. In particular, on a percentage basis, which topic areas do the UK champion, and which are lagging behind? And on a collective basis, are there topic areas that are underfunded globally, that perhaps the UK could try and bolster?

Figure 2 shows this USA dominance of funding. The categories are ordered as before, in descending size based on RCUK activity.

The results provide a reasonable rule of thumb: USA funding is 10 times that of the UK. However, some small, but interesting observations can be found.

1) As a percentage, we spend more on Infection, Neurological, Inflammatory, Mental Health, Reproduction, and, oddly, even as a total, more on 'other' (although 'other' is such an odd category perhaps nothing can be gleaned from this insight)

2) The NIH, as a percentage, spend more on Cancer, Cardiovascular, Metabolic, Respiratory, Renal and, enormously, on Injury and Accidents. In the last 5 years only $7.5M was spent in the UK on Accidents, compared to the USA's $296m.

3) Despite the huge difference in overall spending powers, the allocations (excepting Other and Injury) are quite similar:

Figure 2
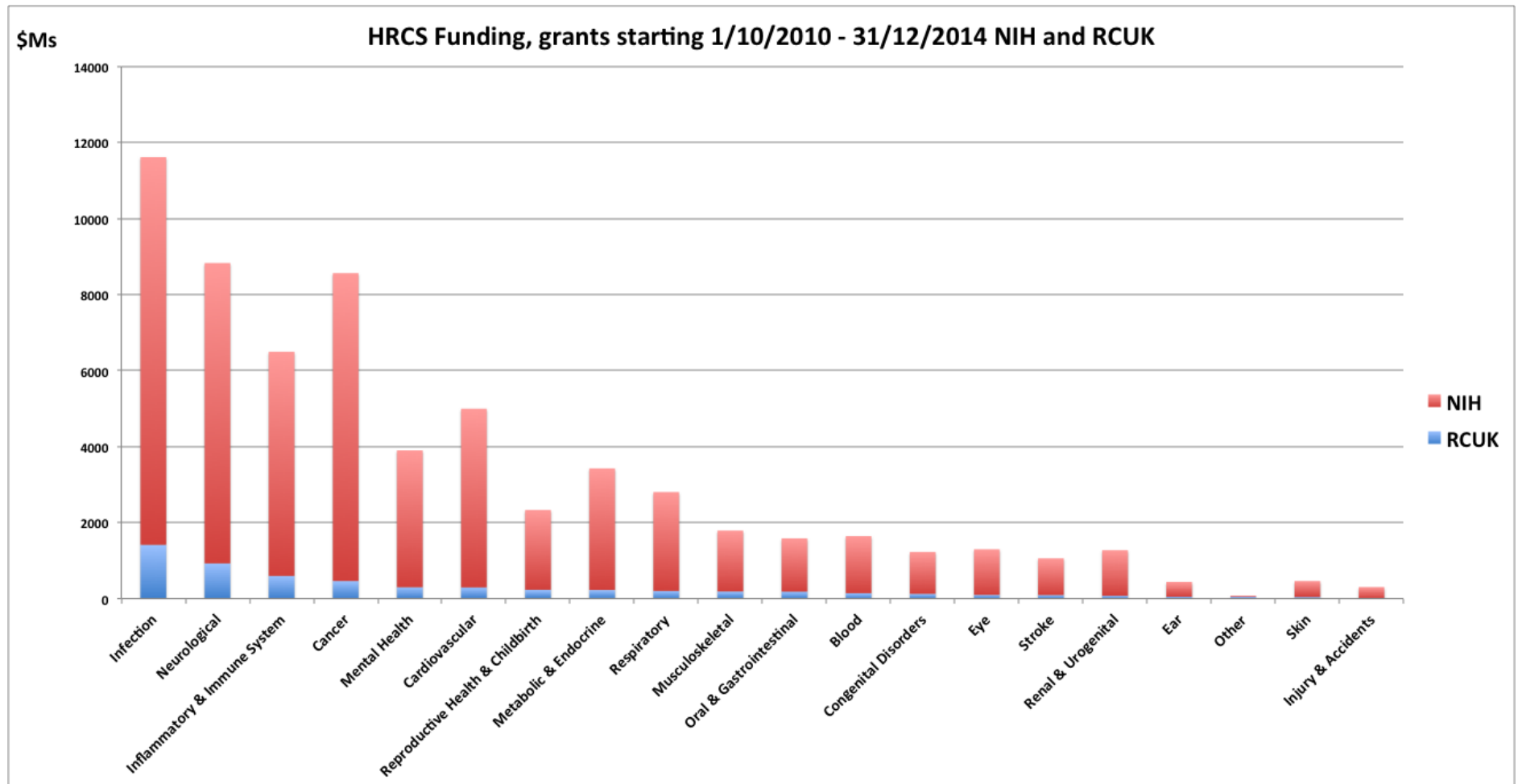
Consideration over what is 'research' and 'biomedical'?

When we look at, say, Mental Health activity by the AHRC we see 32 grants at nearly £6m. However, some of these are historical (and this is true of Wellcome Trust History of Medicine activities which fall into many different topic areas), and perhaps should be excluded. Others are clearly relevant, but are they research? As an example:-

 "AH/L014335/1. Caregivers perceptions of the value of the arts in therapeutic and clinical interventions"

Then there is activity in in, say NERC, in Infection that is not related to human health. For example:-

"NE/M000591/1. Identifying the genetic mechanisms facilitating host range and virulence of a viral pathogen that threatens European amphibian biodiversity"

This is clearly about infectious disease, but concerns amphibians.

Some codes are inappropriate, and ÜberResearch will work on ways of how to remove these without interfering with the solid results in biomedical funders. For example, this EPSRC grant was coded to 'Infection':-

"EP/M001946/1. Characterisation of electron transport in bacterial nano-wire proteins through high performance computing and experimentation"

The expression of bacterial and microbial activity has lead the algorithm to consider this to be Infection (as most bacterial studies are about underlying infections).

Although it appears that the majority of returns in funders outside of Biomedical research are worth investigating, there are more false-positives in these Funders than within Biomedical funders. This is not surprising, as the semantic categories are predicated on biomedical application. ÜberResearch will look at creating a pre-filter to try and identify biomedical activity before applying HRCS codes to non-biomedical funders. However, these insights are still useful for a broad-brush indicative overview, which cannot be undertaken with manual tagging.

Conclusions

1) There are a great many grants from funders outside of traditional biomedical funding that do have relevant activity that should be considered when looking at total research affecting each topic area.

2) There needs to be some discussion as to when an activity is relevant or not. Does it have to be 'research'? Is it only about human and domesticated animal health and welfare?

3) Once a decision is made about the question of relevance of non-research and non-human activity in terms of reporting of HRCS activity ÜberResearch needs to investigate ways of following those guidelines.

4) Based on the data feedback the HRAF can provide to ÜberResearch they will improve the HRCS coding system so that consideration can be given as to how to utilise this system for the review in, and running up to, 2020.