Meaningfully assessing effectiveness of interventions aimed at reducing health inequalities through tackling social determinants of health is challenging. Complex systems and real world delivery environments render many traditional academic tools obsolete. Does the production of big data and possibilities this opens up in the realm of data science offer potential solutions?

**Methods:** 1) Accessing large datasets from across the architecture of health and social services, including primary and secondary care, police, education and linking these using anonymised unique identifiers. 2) Machine learning (using algorithms developed in R Studio) can be employed to identify associations between grouped diverse indicators across the lifecourse. 3) Probabilistic models can then be developed to support better design, targeting and measurement of service redesign and implementation of interventions. 4) Embedded action learning and ethnographic approaches can be used to provide useful practical support to development of interventions and to provide quantitative and qualitative data to allow triangulation with patterns identified through the data science techniques. 5) Frontline staff and the general public will contribute to the refinement of specific research questions to be asked of the data.

**Results:** Currently, identification of the range of datasets and data quality held by public sector actors in Scotland is ongoing. This will be followed by establishment of formal information governance arrangements before algorithms will be trained on training data and applied to the full range of datasets identified, linked and made available to the research team.

**Discussion:** The experience of the Links Worker Programme, which implemented a new social practitioner role within GP practices in deprived neighbourhoods in Glasgow, highlighted the difficulty in sensitively measuring the biological impact of socially oriented interventions. Advances in big data and data science may or may not ultimately aid enhanced understanding of social determinants of health and support reduction in health inequality, certainly though, it is a prescient and logical area to investigate and the inherent potential is exciting. The two pronged approach of data science and action learning supporting robust record keeping on the ground, may offer viable and ‘robust enough’ alternatives to RCTs in real world environments.

**Conclusion:** By time of conference the first findings from analysis of large datasets linking Child Health and Education data across Scottish local authorities will be available.

**Lessons learned:** Current procedures for identifying and accessing publically held data are cumbersome and time consuming.
Limitations: The extent, nature and completeness of the datasets will to some extent guide the research.

Future research: Future collaborations with user experience and user led design will be of importance in helping design end user interfaces that are aligned to provide data amenable to data science approaches, should these prove useful in the field of social determinants/health inequalities.

Keywords: data science; health inequity; social determinants; service enhancement