Sharing and linking police data with health data for public benefit:

Aims: To assess the feasibility of linking police data with health data from hospitals, Accident and Emergency (A&E) Departments and General Practitioners (GPs). Linking police data to healthcare data is not straightforward. Firstly, the data in police systems are often in a narrative format that contain highly sensitive and disclosive information. Although text data is rich descriptively, anonymising and analysing these data is challenging. Secondly, police services in different parts of the country use different software systems, which makes harmonising data across all different police forces and other agencies very difficult. Thirdly, linking entire police datasets with healthcare data has not been previously attempted and it is possible that the risks might appear greater than any benefits.

The overarching aim was to assess whether the benefits from linking police and healthcare data outweigh the effort required and any risks involved. The pilot therefore sought to determine, first, whether text mining could be used to code the text data fields in the Domestic Abuse, Stalking and Harassment (DASH) risk assessment on police data systems. Second, what the views and opinions were of police staff on using a single system and on sharing data with other agencies. Finally, to examine what learning was feasible from linking the DASH assessment with health data. The findings were presented at two workshops at which recommendations for next steps were discussed.

Methods and findings: Text mining methods were used to extract data from the "summary of incident" field, the name of the agency that a PPN referred to, or agencies already involved with the family. This work found that these methods would be able to correctly identify 18 agencies, and these could be coded within the IT department of a police service, enabling textual data to be transformed into a coded format that could be subsequently anonymised and shared. This method could also be used to identify other data of interest.

Interviews with 36 individuals from the four police services in Wales, representatives of the Violence Prevention Unit and Police Liaison Unit, identified that the Niche system was the most used and was highly regarded. Interviewees felt that it would be a positive development if all forces moved to the same system. However, work is still required to ensure that there is a minimum dataset for each area as there is high variability in data quality on the systems. The main barrier prohibiting sharing data is a lack of knowledge on what was possible to share. However, this could be overcome by an unambiguous framework endorsed at a high level of what data should be shared and what data should not be shared.

The exemplar case study linked PPN data with GP and hospital data for 8,709 people. This work identified that those who had with an emergency hospital admission/A&E visit/death within 12 months of their PPN referral were those who already had a history of multiple health records such as had previous emergency hospital admissions before the first PPN visit, had prescriptions for painkillers and antibiotics from the GP in the year before the PPN visit, and the perpetrator was more likely to have a criminal history. Analysis also found that pregnant women in a household of a PPN referral were twice as likely to have a low-birth-weight baby (preterm or failure to grow) compared to pregnant women in households without a PPN referral (after adjusting for confounders such as mum age, mum smoker, deprivation). Younger victims (aged 0-20) are at higher risk of having an emergency A&E attendance or emergency hospital admission in the year after the PPN.

Recommendations: Workshops with a combined attendance of 100 people including those from academia, Welsh and English police services, Welsh Government, Home Office, and the third sector concluded that future directions should focus on quick wins and build on showing the benefits of data sharing. There needs to be consultation with the public to ensure public trust in any data sharing proposed. Future work should be undertaken to agree what information should be shared, agree standards for this data and in what format it should be collected and clarity around how data can be shared while meeting data protection regulations.

Conclusions: This work demonstrated the main barriers to sharing are not physical barriers, for example, there is support to use the same data software systems across police forces and to define a minimum dataset for data collection, it is possible to use text mining to extract data and it is possible to link data across agencies and produce findings that can be used to improve service (for example, including a question on previous emergency admission to hospital in the DASH questionnaire could improve identification of high risk individuals). The main barrier to sharing datasets across agencies is a lack of clarity around data governance and what is appropriate to share. Data sharing would require high level support and unambiguous guidelines as to what data can be shared and in what format.