Atrial fibrillation and electronic healthcare records in UK Biobank: Relevance of ascertainment in primary and secondary care for risk factor associations and patient outcomes

A. Von Ende1, C.F. Camm1, S. Parish1, B. Casadei2, J.C. Hopewell1

1University of Oxford, Nuffield Department of Population Health, Oxford, United Kingdom of Great Britain & Northern Ireland
2University of Oxford, Radcliffe Department of Medicine, Oxford, United Kingdom of Great Britain & Northern Ireland

Funding Acknowledgements: Type of funding sources: Foundation. Main funding source(s): British Heart Foundation. Research is undertaken under UK Biobank application 14568.

Background: Electronic health data have played a pivotal role in the development and success of large-scale biobanks, and have enabled a range of epidemiological studies with wide-ranging applications[1]. Hospital admissions (HA) data are often used to identify individuals with disease, but this may result in incomplete ascertainment for conditions such as atrial fibrillation (AF), which often does not require hospitalisation. Understanding ascertainment through primary care (PC) data is essential to determine the potential benefits and drawbacks of using difference data sources for AF ascertainment.

Purpose: To explore potential differences between AF cases identified from PC and HA data with respect to participant characteristics, timing of ascertainment, and AF-related sequelae.

Methods: UK Biobank is a large prospective study with PC and HA data available in 230,000 participants[2]. Incident AF cases were defined using a combination of Read clinical codes in PC data and ICD-10 diagnostic and OPCS-4 procedural codes in HA data. Individuals were divided into 3 groups representing whether AF was defined using PC records only (PC-only), hospital admissions only (HA-only), or both (PC+HA). Descriptive statistics and multinomial regression were used to describe differences between groups in terms of ascertainment timing, risk factor associations, and post-AF stroke and death.

Results: During a median follow-up of 7 years, 7,142 incident AF cases were identified: 22% through PC-only (1571 new cases), 30% through HA-only, and 48% via PC+HA. In the latter group, when AF was first identified through PC (3827), there was an average lag of 1.3 years before it was identified in HA data, but a 5-year lag in the opposite direction (Figure 1).

The patterns of associations for baseline age, sex, BMI, smoking, blood pressure, and composite clinical risk (CHARGE-AF[3]) suggested little evidence of systematic differences between the groups (Figure 2). However, the associations of AF with baseline diabetes, heart failure, myocardial infarction, and an AF polygenic risk score[4] suggested HA-only cases had a higher prevalence of cardiometabolic comorbidities, but lower overall genetic AF risk.

Moreover, HA-only AF cases had similar rates of subsequent stroke but substantially higher rates of death (73.2 per 1000 person-years) compared to the PC-only and PC+HA groups (6.6 and 23.1, respectively).

Conclusion: PC data identifies 28% additional AF cases than HA alone in UK Biobank, with important implications for the power of epidemiological studies. The strength of associations between risk factors and AF varies by ascertainment source, as do the rates of AF-related sequelae, suggesting that further studies comparing treatment in different groups may be warranted. Furthermore, for AF identified in HA alone, improved consistency and timeliness of routine recording of AF in PC may have important implications for management and risk of sequelae.
Time from primary care AF record to AF hospital admission record, and vice versa

Inset plots represent proportion ascertained within 3m, 6m, and 1 year. Analyses exclude participants who had a prior ascertainment in the alternate source.
Association of baseline AF risk factors with atrial fibrillation, by source of ascertainment

- **Age (5 yrs)**: $P_{het} = 0.001$
- **Male sex**: $P_{het} < 0.001$
- **BMI (5 kg/m2)**: $P_{het} < 0.001$
- **SBP (20 mmHg)**: $P_{het} = 0.001$
- **Current smoker**: $P_{het} < 0.001$
- **Hypertension**: $P_{het} = 0.002$
- **Diabetes**: $P_{het} < 0.001$
- **Heart failure**: $P_{het} = 0.004$
- **Myocardial Infarction**: $P_{het} = 0.007$
- **CHARGE-AF score (SD)**: $P_{het} < 0.001$
- **Polygenic risk score (SD)**: $P_{het} < 0.001$

Based on multinomial logistic models with 'No AF' as the reference category.

PC-only; primary care only; PC+HA, primary care and hospital admissions data; HA-only, hospital admissions data only; $P_{het}$, $P$-value for test of heterogeneity.