BMJ Open

UNDERSTANDING REPEATED NON-ATTENDANCE IN HEALTH SERVICES - PILOT ANALYSIS OF ADMINISTRATIVE DATA AND FULL STUDY PROTOCOL FOR A NATIONAL RETROSPECTIVE COHORT

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-014120
Article Type:	Protocol
Date Submitted by the Author:	01-Sep-2016
Complete List of Authors:	Williamson, Andrea; University of Glasgow, GPPC, School of Medicine, Dentistry and Nursing, MVLS Ellis, David; University of Lancaster Wilson, Philip; University of Aberdeen McQueenie, Ross; University of Glasgow, GPPC, School of Medicine, Dentistry and Nursing, MVLS McConnachie, Alex; University of Glasgow, Robertson Centre for Biostatistics
Primary Subject Heading:	Public health
Secondary Subject Heading:	General practice / Family practice, Epidemiology, Health services research
Keywords:	missed appointments, data linkage, administrative data, health utilisation, health inequalities, social vulnerability

UNDERSTANDING REPEATED NON-ATTENDANCE IN HEALTH SERVICES - PILOT ANALYSIS OF ADMINISTRATIVE DATA AND FULL STUDY PROTOCOL FOR A NATIONAL RETROSPECTIVE COHORT

Andrea E Williamson (corresponding author), General Practice and Primary Care, School of Medicine, Dentistry and Nursing, MVLS, University of Glasgow. andrea.williamson@glasgow.ac.uk Tel: 0141 330 8330

David A Ellis, Department of Psychology, University of Lancaster. d.a.ellis@lancaster.ac.uk

Philip Wilson, Centre for Rural Health, Institute of Applied Health Sciences University of Aberdeen. P.wilson@abdn.ac.uk

Ross McQueenie, General Practice and Primary Care, School of Medicine, Dentistry and Nursing, MVLS, University of Glasgow. Ross.mcQueenie@glasgow.ac.uk

Alex McConnachie, Robertson Centre for Biostatistics, Institute of Health and Wellbeing, MVLS University of Glasgow. Alex.McConnachie@glasgow.ac.uk

Word count 4284

ABSTRACT

Introduction

Understanding the causes of low engagement in health care is a prerequisite for improving health services' contribution to tackling health inequalities. Low engagement includes missing health care appointments. Serially (having a pattern of) missing general practice appointments may provide a risk marker for vulnerability and poorer health outcomes.

Methods and analysis

A proof of concept pilot utilising GP appointment data and a focus group with GPs informed the development of missed appointment categories: patients can be classified based on the number of appointments missed each year. The full study, using a retrospective cohort design, will link routine health service and education data to determine the relationship between general practice appointment attendance, health outcomes, health care utilization, preventive health activity, and social circumstances taking a life course approach and using data from the whole journey in NHS health care. 172 practices will be recruited (approximately 900,000 patients) across Scotland. The statistical analysis will focus on two key areas; factors that predict patients who serially miss appointments, and serial missed appointments as a predictor of future patient outcomes. Regression models will help understand how missed appointment patterns are associated with patient and practice characteristics.

We shall identify key factors associated with serial missed appointments and potential interactions that might predict them. A better understanding of these may also help inform future health promoting care across the health system.

Ethics and dissemination

The results of the project will inform debates concerning how best to reduce non-attendance and increase patient engagement within health care systems. Significant non-academic beneficiaries include governments, policy-makers and medical practitioners. Results will be disseminated via a combination of academic outputs (papers, conferences), social media, and through collaborative public health/policy fora.

STRENGTHS AND LIMITATIONS

- Important question relating to health service component of tackling health inequalities
- Power of a large dataset following patients' journey across the whole health care system
- Utilising data security and linkage capabilities in a sensitive and robust manner
- Having a clear yet flexible data analysis plan utilising the expertise of a multi-disciplinary research team
- Limitations of using administrative data from a range of data sources of variable data quality.

KEYWORDS

Missed appointments, data linkage, administrative data, primary care, health utilisation, health promotion, health inequalities, social vulnerability

INTRODUCTION

Tackling health inequalities is a global health priority[1] and for health service provision to have an effective role, we should understand better the reasons behind, risks associated with, and needs of patients who do not engage effectively with health care provision (even if it is free at the point of access); and tailor services better to meet those needs. There remains a lack of published work concerning repeated missed appointments, but previous research typically focuses on the financial costs associated with non-attendance. One estimate has placed the cost of missed UK general practice (community based family medicine) appointments at £150 million per year[2]. More recent Scottish government data suggest that each missed hospital outpatient appointment costs National Health Services (NHS) Scotland £120[3]. International data on costs to health care systems is sparse. In a complex adaptive system such as health care, the financial costs are contestable because clinicians will 'catch up' or get on with other care or administrative tasks. What is important are the costs of opportunities missed for improving patients' health and the potential for substantial long-term savings to health systems'.

To date studies investigating missed appointments have focused on single missed appointments or single disease outcomes and have indicated they are associated with poorer health outcomes[3-6]. Studies of single missed appointments have produced conflicting results when it comes to designing effective interventions

that can increase attendance[7-10]. This may be due to a reliance on small samples in disparate settings[11-15] and conflation of patients who occasionally miss appointments with patients who have an established pattern of missing many.

The Health and Social Care Information Centre in England has recently published data about repeated missed appointments. From their analysis of recorded missed outpatient hospital appointments in England one in 50 patients (65,590 of 3.5 million) who missed an appointment failed to attend three or more further appointments within three months[16].

We hypothesise that repeated missed appointments reflect a pattern of behaviour. We use the term 'serially' missing appointments to reflect this pattern, which may be interrupted by attended appointments. Clinicians do report that patients who serially miss appointments are of particular concern because they may have very poor health, may be socially disadvantaged and are high users of unscheduled care compared to patients who occasionally or never miss appointments[17].

There is accumulating evidence that negative experiences in early life have pervasive consequences for health over the life course including 'extensive evidence of a strong link between early adversity and a wide range of health-threatening behaviours'[18]. This body of work therefore provides a conceptual framework for better understanding 'chaotic lives'[19] as an explanatory factor in health utilization behaviours such as missed appointments, and introduces the possibility that serial missed appointments contribute to the inverse care law, which states that health care provision is least likely to be provided to those that need it most[20].

In the UK publicly funded general practice (GP) provides almost universal coverage for the population and generates around 90% of health contacts. Appointment making is typically under the control of each patient directly. General practice appointments are therefore a sensible starting point for this study of health and other outcomes across patients' life course, and have relevance for global health systems where patients have direct access to a wider range of health care specialties.

The overarching study question is: is serially missing GP appointments a risk marker for vulnerability and poorer health outcomes and thus a useful target for developing interventions to improve engagement in health promoting care across the health system?

Scotland has an established data linkage infrastructure which is under continuous development. This pathfinder study will for the first time link large general practice datasets with data from across patients' whole journey through health care. We describe the aims and research questions for this study, the pilot work that was undertaken to inform it, and the resultant research protocol for the full study based in GP practices in Scotland.

METHODS AND ANALYSIS

Aim and Research questions

The overall aim of the study is to determine the relationship between general practice appointment attendance, health care utilization, preventive health activity, health outcomes, and social circumstances taking a life course approach and using extracted health service and other relevant administrative data.

A pilot study sought to answer the first research question described below. The subsequent questions underpin the full research protocol which compares cohorts of Scottish patients (from birth to older people) who never, occasionally and serially miss GP appointments.

Figure 1: Study research questions

Pilot study

The pilot study was separated into 2 sub-sections: a 'proof of concept' quantitative data analysis and a focus group to inform and refine definition development (research question 1).

Proof of concept

Research that uses GP appointment data has not previously been conducted using the clinical recording systems in the Scottish NHS. A proof of concept pilot study was undertaken utilising the NHS Trusted Third Party (TTP) Albasoft with 67,705 patient records to determine whether retrieving appointment data was feasible, to refine other data parameters, and to inform the definition development as described in research question 1. An additional confidentiality control means that the research team do not know the identity of the recruited GP practices.

Additional file 1 describes the definition and role of TTPs.

Albasoft purposively recruited 10 Scottish practices on our behalf with the following characteristics:

Figure 2: Pilot practice recruitment

Focus group

A focus group was conducted in September 2015 with five GP participants purposively sampled to represent a range of perspectives. Additional file 2 describes their characteristics. AEW conducted the focus group and the analysis. DE attended the focus group and presented pilot quantitative data for discussion. Additional file 3 describes the topics covered.

Focus group results

Participants reported making clear distinctions between patients who occasionally miss appointments and those who miss many. Patients who occasionally miss appointments do so because a crisis or another understandable circumstance has arisen; patients who serially miss appointments, described as missing more than two or three appointments (SMA) can be easily identified by GPs.

They were described as tending to have mental health, addiction, and/or social issues. They were described as high risk or vulnerable with concerns about their wider family. Patients who SMA were viewed as being different from the general GP population and being more likely to have 'chaotic' lifestyles associated with housing instability, money problems, a 'panicked lifestyle'. Patients who SMA were also described as being unable to manage GPs' expectation of them and fit into GPs' pre-determined slots. On the other hand it was said that not all patients who SMA can be viewed as high risk; some patients do not value free health care. It was reported that some patients who SMA go on to book another appointment the next day.

Missed appointments were viewed as being more prevalent in practices in deprived settings, but occurred in affluent areas too. In the affluent setting they were

important for a minority of marginalised, isolated patients who were viewed as living 'chaotic' lives.

Practices do not have protocols for managing patients who serially miss appointments (SMA) because response is dependent on the patient's context. GPs understood that SMAs usually mean patients with complex needs with workload implications for the practice. Strategies described were varied, including allowing patients only to book on the day, seeing the patient when they walk in, or the GP booking the follow up appointment for the patient- a relationship building strategy. This could still lead to patients missing an appointment, even just a couple of hours after it was made. Some practices do remove patients from their list for SMA and this created tension with other practices. Some practices have a negative view of patients who SMA.

Figure 3 Focus group recommendations for the full study design

Results- definition of serial missed appointments

Following the pilot analysis, data were cleaned and appointment rules applied to categorise appointments as attended or missed (DNA). Appendix 4 describes this process. This was primarily based on the 'in' and 'out' time recorded for that appointment. If this was recorded as '0' then the appointment was classified as Did Not Attend (DNA). For each patient the total number of appointments made during the three-year period was calculated as well as the number and percentage of appointments missed. Appointments that were recorded incorrectly in the system were removed at this stage. This included appointments where

administrative records had remained open for over 24 hours, making it difficult to confirm that these were genuine appointments and not administrative errors. The pilot appointment rules are set out in table 1 below.

Data description	Reason for removal
total appointment time < 0 min	Record open for more than 24 hours
total waiting time < 0 min	Record open for more than 24 hours
appointment <2 min	Not a medical appointment
administrator slot	Not a medical appointment

Table 1 Rules to identify genuine appointments

A pilot analysis of 67,705 patient records showed that while just over 60% of our sample missed no appointments, over 30% missed one or more appointment during the three-year period with nearly 10% of patients missing three or more appointments. Assuming that our final sample provides a similar distribution, we will classify patients based on the number of appointments missed over the last three years as follows:

***Never* missed appointments per year, 0**

***Low* missed appointments per year, <1**

***Medium* missed appointments per year, 1-2**

***High* missed appointments per year, 2 or more**

FULL STUDY PROTOCOL

Recruitment

Our target recruitment is 172 GP practices from across Scotland ensuring that we had a spread of urban, rural and practices characterised by serving areas of blanket high socio-economic (Deep End) practices. This will provide approximately

900,000 patient records for inclusion in the study. The following is the information request made to practices.

Figure 4 Information request sent to target practices

Data Handling

The TTP is recruiting the practices on our behalf and will undertake some specific data aggregation before transferring the data securely to the National Safehaven for analysis. 'Safe Havens are specialised, secure environments supported by trained, specialist staff where data in electronic patient records can be processed and linked with other health data (and/or non-health-related data) and made available for analysis to facilitate research while protecting patient identity and privacy'[21]. These are: calculating urban rural classification, SIMD decile, categorising ethnicity into 'non BME (Black and Minority Ethnicity)', 'visibly BME', and 'non visible BME' and rounding distance to practice/emergency department to the nearest kilometre for each patient record. Once in a Safehaven, additional steps will be taken to ensure that acceptable anonymization principles are being applied, especially in relation to reporting of sensitive social vulnerability data and reporting of rare conditions.

A new data file containing the appointment history for each patient record will be generated, which will be merged with individual patient information (Additional file 4 sets this out based on our pilot data set)

Data Linkage

This will be conducted as access permissions and data sets become available for linkage so will be incremental. Each administrative data source is available for different time periods (e.g. hospital inpatients since 1981 and education outcomes since 2002) and this will be made explicit when interpreting the results. The TTP will provide the Safe Haven indexing team a file containing the GP dataset Community Health Index (CHI) number and other patient identifiers. This forms the cohort for the study. All data providers will provide identifiers to be probability matched to the study cohort by the Safehaven linkage team, who will return a set of unique index numbers for those individuals successfully matched to the study cohort; each data provider will receive a different set of unique index numbers, and will use these index numbers as the basis of their data extract. Each data extract will be submitted to the Safehaven linkage team, who will replace the different index numbers with a common number across all files. This common number is the unique patient identifier that the research team will work from to analyse the linked data.

Figure 5 Proposed data sets for linkage with GP data

Appointment validation and categorisation

Each appointment will be coded by the session type recorded by the practice (eg book on day appointments, or immunization clinic) and then further by professional type (eg GP partner, practice nurse). These descriptions are determined by individual practices so categorisation will be conducted by the GPs in the research team. The appointment rules set out in the pilot study will be

applied. A sensitivity analysis based on the time the appointment takes will then also be conducted by comparing a random sample of patient appointments according to

Figure 6 Random sample of GP appointments for validation and sensitivity analysis

The appointment rules will be refined based on this. The time interval cut-off for apparently attended appointments will be determined by utilising the time interval that most accurately matches to actual attended appointments. Slots designated non face to face appointments will then be removed leaving only attended and non- attended face- to- face appointments. The appointment categories described from the pilot study regarding non- attendance for all patients will then be applied to the yearly average number of missed appointments over the three year extract period to generate the four categories of patients for further analysis. Using an average over three years takes account of what is recognised in the frequent attenders (rather than non- attenders) literature- that patients' appointment behaviour may vary over time in relation to illness episodes or social crises[22].

Accounting for patient turnover

This study seeks to ensure the inclusion of patients who are marginalised and who are often missing from health service studies. There is evidence of overlap between patients who miss appointments and those who are removed from practice lists[23], a recognition of the impact that geographical boundary areas has on patients who move around[24]; notwithstanding the gap in the literature about

registration interruptions for patients who may go to prison or patients who remain unregistered once they are removed from GP practice lists. We will therefore summarise the numbers of patients joining and/or leaving their practice during the study period; with reasons where this information is available. We will seek to provide a full analysis of the data available for these patients and compare these with the patients who are registered for the 3 year study period.

Statistical analysis

Our statistical analysis is based on the study being a retrospective cohort study. We will focus on two key areas; predictors of high rates of serial missed appointments, and serial missed appointments as a predictor of future patient outcomes.

Patient characteristics and practice characteristics may be associated with high rates of serial missed appointments. Analyses will initially be descriptive, summarising the rate of missed appointments in relation to the other factors recorded at the point of entry to the study. Associations with patient characteristics will be assessed as a whole, and in relation to different types of practices (e.g. separately in rural and urban practices).Subsequently, we will use regression models[25] to help us understand how our categories of missed appointments are associated with patient and practice characteristics. Specifically, Poisson or Negative Binomial regression will be used.

When considering other outcomes in relation to serial missed appointments, the missed appointment rate category (none, <1, 1-2, or >2 per year) will be considered as the predictor. Appropriate regression models, according to the outcome variable, will be used to assess whether any associations with serial

missed appointment rates are independent of other patient- or practice-level factors.

We also plan to measure the factors recorded during the study interval associated with having a lot of missed appointments. We will explore whether these differ from the predictive factors already recorded at entry to the study.

Potential Predictors of frequent non- attendance

Demographics

Patients' age, gender, minority ethnic group status (where available), deprivation decile, rural/urban split, number of address moves, distance lived from GP practice and distance from nearest A&E will be explored.

Health conditions

Health conditions will be reported using four separate variables. Firstly by the incidence of multi-morbidity calculated from extracted Read codes based on previous counts in Scotland[26], secondly descriptions of health conditions based on the priority 1 Read codes that GP practices in Scotland use to populate patients' key information summaries (KIS) for GP out of hours services. This is novel work as a coding structure has not previously been applied to these Read codes. Thirdly, a count of psychotropic medicine prescriptions based on the British National Formulary will be generated. This is in order to describe levels of psychological morbidity that are not captured by diagnostic criteria. These three variables will then be compared to the ICD 10 coding data from patients' secondary care linked data compiled from hospital admissions and outpatient attendances. Diagnostic

data from emergency department attendance was deemed not of sufficient quality to utilise.

Social Vulnerability

One aspect of this study which is particularly ground-breaking is our investigation of retrievable information about patients' social vulnerability. The Adverse Childhood Experiences (ACE) questionnaire[27] will be utilised as a template to match its nine descriptors of adversity to relevant Read codes in the patient's GP record. In addition, coding that maps the consequences of ACE will be analysed. A recent quantitative evaluation of Severe and Multiple Disadvantage will also be matched to GP Read codes. This examines the overlap of patients being homeless, in substance misuse services, or in prison over the preceding year[28]. Further, an exploration of additional Read codes that describe social vulnerability will be mapped. An anonymised text search linked to Read codes from the dataset will provide additional information about social vulnerability as it is recorded in the free text portion of GP records. Both of these taken together will provide the first research evidence about the breadth and depth of social vulnerability recording by GPs.

Health screening and utilisation

Read coding in relation to cervical, breast and bowel screening attendance will be retrieved in addition to the proportion of patients who have had their blood pressure checked, have participated in child health surveillance and vaccination programmes across the life course. A sub-analysis of utilisation of practice nurse and other health care professional's' appointments in the practice will also be conducted and include an exploration of the relationship between attending all

primary care appointments and categories of non- attendance. This is because data from the GP focus group suggested there is overlap between patients who are serial non-attenders with patients who are very frequent attenders. We will therefore consider the rate of attending appointments as a potential predictor of the rate of non-attendance. Referrals that GPs make into other primary and secondary care services will also be analysed. Outpatient attendances, hospital admissions and utilisation of emergency departments, NHS 24 triage, GP out of hours, and ambulance services will also be conducted when linked data become available with a specific focus on how this relates to unmet need, for example how GP appointment category relates to patterns of other health care utilisation between scheduled and unscheduled secondary care use.

Engagement with healthcare

An analysis of GP Read codes and linked secondary care data that relate to patients not attending primary and secondary care appointments, refusing screening, being exception-reported (ie excluded from the denominator population) from the Quality and Outcomes Framework (QOF) system for performance measurement in general practice, practices' measures of non-engagement with care for long term conditions, taking 'irregular discharge' from hospital (when patients leave against medical advice), and not waiting to be seen in emergency departments will be conducted.

Family linkage

Diagnoses of children who are able to be linked through family linkage will be analysed by their mother's appointment category. This is contingent on the child being included in the practice study population.

Education data

Attendance at school, exclusion from school, and educational attainment when leaving school will be made with approximately a sixth of our patient cohort for whom linked education data is available. This has potential to inform future planning about interventions to reduce serial missed appointments.

Practice level data

Each patient record will be allocated a unique practice ID enabling the research team to analyse each patient record output clustered by practice. This will be proportion of patients aged over 75, by ethnicity(proportion BME), patient rurality, patient level of deprivation decile, patient distance to practice, distance to A&E appointments offered/engaged, days from when appointment is made, multi-morbidity count, ACE score more than 4, Severe and Multiple Disadvantage score, hospital referrals, and proportion of each appointment category by practice. These analyses and output will be refined as the study proceeds taking patient level findings and multilevel modelling that takes the interaction between the patient and the practice into account.

Health outcomes

Mortality data regarding date and cause of death will be utilised both from the GP and linked data and from linked obstetric outcomes (from the Scottish Birth Record) for relevant women.

ETHICS AND DISSEMINATION

This pathfinder linkage retrospective cohort study is necessarily complex in design and implementation because although cross-sectional it seeks to take a life course

approach and follow the patients' journey through health care. Careful attention and significant resource has been devoted to the consideration of patient privacy and confidentiality. This has been integrated throughout the design of the study alongside the necessary data access and handling permissions. Additionally a study of this nature, which involves stakeholders across the NHS and other public services, requires a flexible time frame to allow access to raw data and to share findings between members of the research team based in several institutions. The proof of concept pilot did not require ethical permission because it was considered service evaluation with the agreement we would not publish results about the practices which took part; ethical permission to conduct the GP focus group and publish the results was obtained by the MVLS ethics committee, University of Glasgow (ref 200140181). A letter of comfort was obtained from the West of Scotland NHS ethics committee and the MVLS ethics committee that the full study did not need health service ethics permissions. Multi- site NHS R&D approval for the full study was obtained for all Scottish Health Boards (NRS16/186358).

Due to the sensitive nature of administrative data from the NHS and public education system in Scotland, the datasets generated and/or analysed during the current study will not be publicly available. They have been made available to the research team under controlled access conditions and strictly for the purposes of this research study only. Summary data at the level of disclosure checked output from the National Safehaven and statistical code can be requested from the corresponding author on reasonable request.

Planned output

Alongside peer reviewed academic papers reporting the findings described above, the following are planned.

Data Visualisation

Several web pages will be built to sit alongside key results. This will allow for the rapid construction of interactive data visualisations which will be created using "Shiny"[29], a web application framework for R. A simple platform will allow researchers and collaborators to interact with the analyses in real-time and generate custom tables and graphs as required. It can also provide non-experts with access to simple and complex statistical analysis using a point-and-click interface. This will not rely on raw data and will simply pull information from the summary descriptive analyses.

Case Studies

We also intend to use case studies to develop and illustrate our findings throughout the course of all our analyses. For example, we will be able to identify typical patient profiles of those who appear to miss many appointments in a very short period of time and compare these events with short and long-term health outcomes.

Conclusion

We shall identify key factors associated with serial missed appointments ranked in order of importance as described, but given the large sample size we shall also be able to consider potential interactions that might predict serially missed appointments.

Finally, this approach also explores the theory that low engagement with health care should be viewed as a health harming behaviour, and will inform the debate

about tackling health inequalities at the health service delivery level. This will allow us to better understand and develop future interventions to reduce serial missed appointments.

REFERENCES

1. Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health. Geneva: World Health Organisation, 2008.

2. George A, Rubin G. Non-attendance in general practice: a systematic review and its implications for access to primary health care. Family Practice 2003;20:178-84.

3. Campbell K, Millard A, McCartney G, et al. Who is least likely to attend? An analysis of outpatient appointment 'Did Not Attend' (DNA) data in Scotland. Edinburgh: NHS Health Scotland, 2015.

4. Neal RD, Hussain-Gambles M, Allgar VL, et al. Reasons for and consequences of missed appointments in general practice in the UK: questionnaire survey and prospective review of medical records. BMC Family Practice 2005;6(47).

5. Mitchell AJ, Selmes T. Why don't patients attend their appointments? Maintaining engagement with psychiatric services. Advances in Psychiatric Treatment 2007;13:423-34.

6. Ciechanowski PS, Russo J, Katon W, et al. Where is the patient? The association of psychosocial factors and missed primary care appointments in patients with diabetes. General Hospital Psychiatry 2006;28:9-16.

7. Lehmann T, Lehmann A, Olivet M, et al. Missed appointments at a Swiss university outpatient clinic. Public Health 2007;121:790-99.

8. Neilson KM, Faergeman O, Foldspang A, et al. Cardiac rehabilitation: health characteristics and socio-economic status among those who do not attend. European Journal of Public Health 2008;18:479-83.

9. Cashman S, Savageau J, Lemay C, et al. Patient health status and appointment keeping in an urban community health center. Journal of health care for the Poor and Underserved 2004;15:474-88.

10. Masuda Y, Kubo A, Kokase A, et al. Personal features and dropout from diabetic care. Environment, Health Prevention Medicine 2006;11:115-9.

11. Gucciardia E. A systemic review of attrition from diabetes education services: strategies to improve attrition and retention research. Canadian Journal of Diabetes Care 2008;32(5365).

12. Murdock A, Rogers C, Lindsay H, et al. Why do patients not keep their appointments? Prospective study in gastroenterology outpatient clinic. Journal of Royal Society of Medicine 2002;95:284-6.

13. Farley R, Wade T, Birchmore L. Factors influencing attendance at cardiac rehabilitation among coronary heart disease patients. European Journal of Cardiovascular Nursing 2003;2:205-12.

14. Woodward V. Non attendance at a nurse-run asthma clinic. Journal Clinical Nurse 1998;7:380-1.

15. Killaspy H, Banerjee S, King M, et al. Prospective controlled study of psychiatric out-patient non- attendance, Characteristics and outcome. British Journal of Psychiatry 2000;176:160-5.

16. Health and Social Care Information Centre England. Provisional monthly topic of interest: outpatient did not attend (DNA) appointment. London, 2014.

17. Watt G. Deep End Report 19. Access to Specialists. The 19th meeting of 'General Practitioners at the Deep End'. University of Glasgow, 2013.

18. Shonkoff JP, Garner AS, The Committee On Psychosocial Aspects Of Child And Family Health COEC, Adoption, And Dependent Care, And Section On Developmental And Behavioral Pediatrics,. The Lifelong effects of Early Childhood Adversity and Toxic Stress. Pediatrics 2012; 129.

19. Hussain-Gambles M, Neal RD, Dempsey O, et al. Missed appointments in primary care: questionnaire and focus group study of health professionals. Bristish Journal of General Practice 2004;54:108-13.

20. Tudor Hart J. The Inverse Care Law. The Lancet 1971;297:405-12.

21. The Scottish Government. Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics. Edinburgh2015 [Available from: http://www.gov.scot/Publications/2015/11/4783/2.

22. Smits FTM, Brouwer HJ, van Weert HCP, et al. Predictability of persistent frequent attendance: a historic 3-year cohort study. British Journal of General Practice 2009;59:114-19.

23. Williamson AE, Mullen K, Wilson P. Understanding “revolving door” patients in general practice: a qualitative study. BMC Family Practice 2014;15(1):1-11.

24. O'Reilly D, Steele K. The removal of patients who live outside the practice boundary. British Journal of General Practice 2005;55:384-86.

25. Zuur AF, Ieno EN. A protocol for conducting and presenting results of regression-type analyses. Methods in Ecology and Evolution 2016;7(6):636-45.

26. Barnett K, Mercer SMN, Watt G, et al. Epidemiology of Multi-morbidity and implications for health care, research and medical education: a cross sectional study. The Lancet 2012;380(9836):37-43.

27. Bellis MA, Ashton K, Hughes K, et al. Welsh Adverse Childhood Experiences (ACE) Study. Adverse Childhood Experiences and their impact on health-harming behaviours in the Welsh adult population. NHS Wales, 2015.

28. Bramley G, Fitzpatrick S, Edwards J, et al. Hard Edges Mapping Severe and Multiple Disadvantage England. . England: Lankelly Chase Foundation, 2015.

29. Ellis DA, Meridian HL. Thinking outside the box: developing dynamic data visualizations for psychology with Shiny. Frontiers in Psychology 2015;6.

AUTHORS' CONTRIBUTIONS

AEW is principal investigator for the study. DAE, PW and AMcC are co-investigators on the study and RMcQ is the research assistant. AEW conceived and developed the initial research proposal, reviewed the literature, conducted and analysed the pilot focus group, contributed to analysis and interpretation of the quantitative pilot data, developed the predictors, outcomes and associations of interest and led on writing the paper. DAE supported the development of the initial research proposal, reviewed the literature, conducted and analysed the quantitative pilot data, developed the statistical and output plan, and contributed to writing the paper. PW supported the development of the initial research proposal, reviewed the qualitative and quantitative pilot results, reviewed the statistical and output plan and contributed to writing the paper. RMcQ reviewed the statistical and output plan and contributed to writing the paper. AMcC provided expert statistical

input to the study as it was developed, reviewed the statistical and output plan and contributed to writing the paper.

All authors read and approved the final manuscript.

FUNDING STATEMENT

This study is supported by a Scottish Government Chief Scientist Office research grant (CZH/4/41118) with Safehaven and data linkage costs supported in lieu by the DSLS at Scottish Government. These funding bodies had no role in the design of the study, or collection, analysis, and interpretation of data or in writing the manuscript.

COMPETING INTERESTS

The authors declare they have no competing interests.

ACKNOWLEDGEMENTS

We would like to acknowledge all GP practices and GPs who took part in the pilot study. Also colleagues at Scottish Government who are supportive of the study in a variety of ways especially Ellen Lynch in the Health Analytics Division. Dave Kelly's technical and procedural expertise, wisdom and patience as director of our TTP Albasoft Ltd, underpins all of what has been achieved to date.

1. What is a useful definition of never, occasionally and serially missing GP appointments?

2. What are the differences in illness profile, including multi-morbidity across patients' life course between these categories of patients?

3. What are the differences in health service utilization across the primary, secondary, scheduled and unscheduled health services?

4. What are the differences in health outcomes across the whole health system?

5. What are the differences in social vulnerability?

6. Can missed appointments be used to develop a proxy for unmet health need?

7. Can conclusions be drawn to inform rational resource allocation?

8. Is there evidence that supports the future development of targeted interventions to reduce missed appointments?

1. 6 practices in urban and 4 practices in rural settings based on 'rural 8' classification[22] scores
2. 7 of those practices in areas of high deprivation - based on Scottish Index of Multiple Deprivation (SIMD)[23] average patient scores for the registered list
3. 2 practices have high proportions of minority ethnic group patients based on previous work by Albasoft.

1. Participants thought that the most important aspect of the study was to work out whether missed appointments were predominantly a feature of practice behaviour (so the impact of adapted appointment systems that took account of patient behaviour) or a feature of the patients who missed appointments.

2. In terms of practices, participants thought that appointment systems and especially time from booking to appointment date was important.

3. They felt that it was important if SMA was a patient feature to identify the patients whose appointment behaviour could change and those whose could not - as GPs really do this already when they use strategies for managing patient's appointment behaviour.

4. Participants were astonished by the data presented that suggested some patients had missed 25-41 appointments over 3 years and viewed this as 'extreme'. They suggested these data need careful review and postulated it may be a data cleaning issue, an 'anomaly'. They also suggested it might be related to practice factors for example a very tolerant GP; or patient factors; if the data were accurate.

5. Participants also pointed out that the data presented also included patients who serially *attend* appointments. These patients are viewed as having similar characteristics to patients who serially miss appointments and would be a useful additional focus for the study.

6. Participants were surprised that patients who serially missed appointments were more likely to live close to the practice. Participants thought this may be because patients attributed lower value to their appointments because they had to make less effort to attend or it may be that this is a signal that SMA's are predominantly an urban problem as patients in urban areas tend to live closer to their GP practice.

7. The participants cautioned that care needs to be taken that real appointments are captured as they are not always accurately recorded in practice computer systems by appointment type.

8. Participants recommended that we take into account that patients will have clusters of missed appointments when specific events are happening in patient's lives such as a recent major bereavement.

9. The participants also described large variability in practice Read coding for vulnerability.

10. Participants thought it was important that the results of the study be illustrated by case studies of patients as this will be useful for practising clinicians' learning.

The data we are seeking are from the patient records of patients who have scheduled a GP appointment in the 3 years preceding the data extract date:

1. Dates of GP appointments

a. with missed or attended codes, b. session type, c. type of practitioner (e.g. GP, Practice Nurse), d. number of days since appointment made, e. unique practice ID

2. Patient demographic data

a. Age, b. Sex, c. Ethnicity, d. data zone, e. count of patient address moves , f. distance to practice (rounded to km), g. distance to hospital with an emergency department (rounded to km)

3. Clinical and prescribing data, Selected Read codes:

a. priority 1 coding (important health conditions that GPs code for export into the electronic care summary), b. long term condition diagnoses[2] (ref) c. patient vulnerability and adversity factors (ref), d. health screening (breast, bowel, cervical, BP, child health surveillance), e. exception coding (hospital referrals, DNA codes, refused screening, Quality and Outcomes (QOF, payment for targeted long term conditions management) exemption reporting, inappropriate use codes, self- discharge codes), f. specific prescribing information from BNF Central Nervous System (CNS) chapter and additional prescriptions specific to 40 long term diagnoses in b.)

4. Exit codes:

a. patient death, b. patient moved practice.

1. Deaths
2. SMR01 – hospital inpatients & day cases
3. SMR 25 –drug misuse database
4. SMR00 –hospital outpatients
5. SMR04- mental health admissions
6. A&E attendances
7. SMR02 –maternity services including a family index
8. Immunisation records
9. NHS24 – NHS advice help line contacts
10. SAS –Scottish Ambulance Service contacts
11. GP Out of Hours contacts
12. SQA education- attainment and attendance

After appointment category and rules have been applied:

Apparently attended appointments:
100 time interval more than 4 mins
200 less than 4 mins
200 less than 3 mins
200 less than 2 mins excluding 0 time appointments
200 less than 2 mins including 0 time appointments

100 apparently missed appointments

Additional file 1: Definition and role of TTPs

With the increasing demand for statistical, research and service planning information from primary care records a solution is required to reduce the exposure of patient and clinician information to the requesting organisations to a minimum. The recommended (Information Commissioners Office) method of achieving this is by using a trusted third party (TTP) as an intermediary between organisations, which significantly reduces the number of individuals with access to identifiable information. In this case the TTP's role is to provide the technical skills to extract the required information from the Data Controllers electronic records and process this into a form that is both fit for purpose and complies with principal 3 of the data protection act. This may require the removal/replacement of identifiers (anonymisation /pseudo-anonymisation) or the use of redaction techniques when only statistical information is required prior to release of information to the beneficiary.

A TTP is required to operate to strict guidelines as it may only processes data in accordance with instructions from the data controller and to a specification previously agreed by both data requestor and data provider. The TTP acts as a Data Processor on behalf of the Data Controller and abides by the principles defined in the data protection act. It is registered as a data processor with the ICO, provides a secure storage facility which operates procedural, physical and electronic access controls to protect the data it processes and has no specific interest in, not is affiliated to any organisation that has an interest in any data provided. Albasoft maintain a secure data processing and storage facility at the Centre for Health Science adjacent to Raigmore hospital in Inverness, this facility is solely hosted on the NHS network. No information is transferred out with the

NHS network. Its existing middleware platform Escro is an advanced practice based reporting system and is used to securely process data locally at the practice before transferring the results to their secure repository. Albasoft has an established track record as a TTP for the Scottish Therapeutics Utility and increasingly in supporting NHS research. In our study, Albasoft have established data sharing agreements with Scottish GP practices for computerised access to the GP practice data.

Additional file 2: GP focus group participant characteristics

Participant characteristics	Practice setting	Other work roles
• 4 male and 1 female GP • All aged 40-55 years old	• 3 high urban deprivation • 1 urban high affluence • 1 mixed semi-rural with pocket deprivation	• 1 clinical director of a Health and Social Care Partnership • 1 Local Medical Committee member • 1 clinical lead for a national innovation project • 2 with strategic Royal College of General Practitioner roles • 2 members of the 'GPs at the Deep End' steering group

Appendix 3: GP Focus Group Interview Schedule

Introductions:

Name, how long in clinical practice, time in your job, brief description of practice setting.

A priori knowledge/experience of missed appointments

Are missed appointments important? If so why? If not why not?

Can a distinction be made between patients who occasionally miss and those who serially miss GP appointments? If so what are those distinctions? Are they important?

How do you make that distinction in clinical practice? (probe distinctions between individuals and practice settings)

What does it mean for you, your practice and patients? Specifically patients who serially miss?

Present proof of concept provisional data (data cut offs, patient profiles)

What does this data tell us about the issue of serial missed appointments?

What are the obvious things it tells us? What are the surprises? Why?

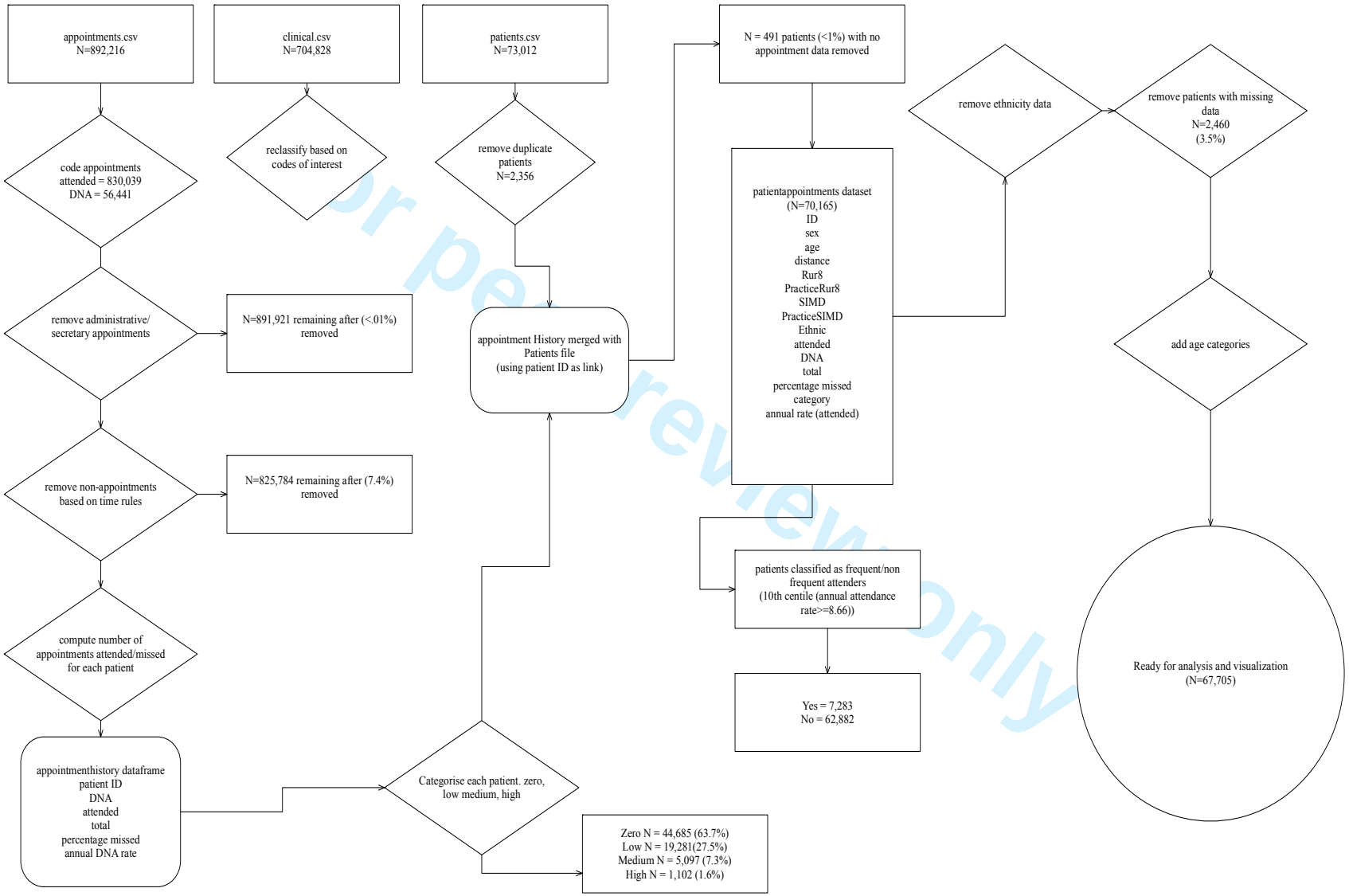
Do you think it misses important aspects of what you think about the issue? Why might that be?

If we present these options about what a definition of a patient who serially missed appointments compared to one who occasionally does, which one do you think is most accurate? Why?

Is there more information that we should look for before deciding we have a definition? What should that be?

Conclusion

Finally, are there aspects of missed appointments and the definition development we have worked on today that we have not yet covered and you would like to tell us about?



BMJ Open

UNDERSTANDING REPEATED NON-ATTENDANCE IN HEALTH SERVICES - PILOT ANALYSIS OF ADMINISTRATIVE DATA AND FULL STUDY PROTOCOL FOR A NATIONAL RETROSPECTIVE COHORT

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-014120.R1
Article Type:	Protocol
Date Submitted by the Author:	14-Dec-2016
Complete List of Authors:	Williamson, Andrea; University of Glasgow, GPPC, School of Medicine, Dentistry and Nursing, MVLS Ellis, David; University of Lancaster Wilson, Philip; University of Aberdeen McQueenie, Ross; University of Glasgow, GPPC, School of Medicine, Dentistry and Nursing, MVLS McConnachie, Alex; University of Glasgow, Robertson Centre for Biostatistics
Primary Subject Heading:	Public health
Secondary Subject Heading:	General practice / Family practice, Epidemiology, Health services research
Keywords:	missed appointments, data linkage, administrative data, health utilisation, health inequalities, social vulnerability

UNDERSTANDING REPEATED NON-ATTENDANCE IN HEALTH SERVICES - PILOT ANALYSIS OF ADMINISTRATIVE DATA AND FULL STUDY PROTOCOL FOR A NATIONAL RETROSPECTIVE COHORT

Andrea E Williamson (corresponding author), General Practice and Primary Care, School of Medicine, Dentistry and Nursing, MVLS, University of Glasgow. andrea.williamson@glasgow.ac.uk Tel: 0141 330 8330

David A Ellis, Department of Psychology, University of Lancaster. d.a.ellis@lancaster.ac.uk

Philip Wilson, Centre for Rural Health, Institute of Applied Health Sciences University of Aberdeen. P.wilson@abdn.ac.uk

Ross McQueenie, General Practice and Primary Care, School of Medicine, Dentistry and Nursing, MVLS, University of Glasgow. Ross.mcQueenie@glasgow.ac.uk

Alex McConnachie, Robertson Centre for Biostatistics, Institute of Health and Wellbeing, MVLS University of Glasgow. Alex.McConnachie@glasgow.ac.uk

Word count 4869 excluding abstract, figures, qualitative data quotes and references.

ABSTRACT

Introduction

Understanding the causes of low engagement in health care is a prerequisite for improving health services' contribution to tackling health inequalities. Low engagement includes missing health care appointments. Serially (having a pattern of) missing general practice appointments may provide a risk marker for vulnerability and poorer health outcomes.

Methods and analysis

A proof of concept pilot utilising general practice (GP) appointment data and a focus group with GPs informed the development of missed appointment categories: patients can be classified based on the number of appointments missed each year. The full study, using a retrospective cohort design, will link routine health service and education data to determine the relationship between general practice appointment attendance, health outcomes, health care utilization, preventive health activity, and social circumstances taking a life course approach and using data from the whole journey in NHS health care. 172 practices will be recruited (approximately 900,000 patients) across Scotland. The statistical analysis will focus on two key areas; factors that predict patients who serially miss appointments, and serial missed appointments as a predictor of future patient outcomes. Regression models will help understand how missed appointment patterns are associated with patient and practice characteristics.

We shall identify key factors associated with serial missed appointments and potential interactions that might predict them.

Ethics and dissemination

The results of the project will inform debates concerning how best to reduce non-attendance and increase patient engagement within health care systems. Significant non-academic beneficiaries include governments, policy-makers and medical practitioners. Results will be disseminated via a combination of academic outputs (papers, conferences), social media, and through collaborative public health/policy fora.

STRENGTHS AND LIMITATIONS

- This study will answer important question relating to the health service component of tackling health inequalities
- A large dataset enables the researchers to follow patients' journey across the whole health care system
- The study utilises data security and linkage capabilities in a sensitive and robust manner
- The study has a clear yet flexible data analysis plan utilising the expertise of a multi-disciplinary research team
- There are limitations of using administrative data from a range of data sources of variable data quality.

KEYWORDS

Missed appointments, data linkage, administrative data, primary care, health utilisation, health promotion, health inequalities, social vulnerability

INTRODUCTION

Tackling health inequalities is a global health priority[1] and for health service provision to have an effective role, we should understand better the reasons behind, risks associated with, and needs of patients who do not engage effectively with health care provision (even if it is free at the point of access); and tailor services better to meet those needs. There remains a lack of published work concerning repeated missed appointments, but previous research typically focuses on the financial costs associated with non-attendance. One estimate has placed the cost of missed United Kingdom (UK) general practice (community based family medicine) appointments at £150 million per year[2]. More recent Scottish government data suggest that each missed hospital outpatient appointment costs National Health Services (NHS) Scotland £120[3]. International data on costs to health care systems are sparse. In a complex adaptive system such as health care, the financial costs are contestable because clinicians will 'catch up' or get on with other care or administrative tasks. What is important are the costs of opportunities missed for improving patients' health and the potential for substantial long-term savings to health systems.

To date studies investigating missed appointments have focused on single missed appointments or single disease areas and have indicated they are associated with poorer health outcomes[3-6]. Studies of single missed appointments have produced conflicting results when it comes to designing effective interventions that can

increase attendance[7-10]. This may be due to a reliance on small samples in disparate settings[11-15] and conflation of patients who occasionally miss appointments with patients who have an established pattern of missing many.

The Health and Social Care Information Centre in England has recently published data about repeated missed appointments. From their analysis of recorded missed outpatient hospital appointments in England one in 50 patients (65,590 of 3.5 million) who missed an appointment failed to attend three or more further appointments within three months[16].

We hypothesise that repeated missed appointments reflect a pattern of behaviour. We use the term 'serially' missing appointments to reflect this pattern, which may be interrupted by attended appointments. Clinicians do report that patients who serially miss appointments are of particular concern because they may have very poor health, may be socially disadvantaged and are high users of unscheduled care compared to patients who occasionally or never miss appointments[17].

There is accumulating evidence that negative experiences in early life have pervasive consequences for health over the life course including 'extensive evidence of a strong link between early adversity and a wide range of health-threatening behaviours'[18]. This body of work therefore provides a conceptual framework for better understanding 'chaotic lives'[19] as an explanatory factor in health utilization behaviours such as missed appointments, and introduces the possibility that serial missed appointments contribute to the inverse care law, which states that health care provision is least likely to be provided to those that need it most[20].

In the UK, publicly funded general practice (GP) provides almost universal coverage for the population and generates around 90% of health contacts.

Appointment making is typically under the control of each patient directly. General practice appointments therefore provide a sensible starting point for this study of health and other outcomes across patients' life course. Subsequent results will also have relevance for global health systems where patients have direct access to a wider range of health care specialties.

Scotland has an established data linkage infrastructure which is under continuous development. This pathfinder study will for the first time link large general practice datasets (including appointment data) with data from across patients' whole journey through health care.

The overarching study question is: is serially missing GP appointments a risk marker for vulnerability and poorer health outcomes and thus a useful target for developing interventions to improve engagement in health promoting care across the health system?

Aim and Research questions

The overall aim of the study is to determine the relationship between general practice appointment attendance, health care utilization, preventive health activity, health outcomes, and social circumstances taking a life course approach and using extracted health service and other relevant administrative data.

A pilot study sought to answer the first research question described below (figure 1). The subsequent questions underpin the full research protocol which compares cohorts of Scottish patients (from birth to older age) who never, occasionally and serially miss GP appointments.

Figure 1: Study research questions

An introduction to the full study protocol is described, followed by the methods and results from a mixed methods pilot study that informed the protocol. A description of protocol participants, data sources, variables and statistical analysis then follows.

METHODS AND ANALYSIS

The full study protocol is for a retrospective cohort study of GP practice patient records linked with secondary care and education administrative records in Scotland.

The study commenced in July 2015 and will finish in December 2017. A pilot study was conducted between July and September 2015 which is described next. The cohort of 909,073 GP patient records for the full study was available in the National Safehaven from September 2016 and analysis of these data is underway. Permissions to access education data is secured, and the outcome of linkage permissions for health data is not yet confirmed

Pilot study

The pilot study was separated into 2 sub-sections: a focus group to inform and refine definition development (research question 1) and a 'proof of concept' quantitative data analysis.

Methods

Focus group

A focus group was conducted in September 2015 with five GP participants. A focus group was judged the most appropriate method to use because we aimed to promote discussion of the topic such that participants would be able to compare

and contrast their own experiences with others from a range of practice and professional experience settings[21]. Linked to this was the aim of asking participants to make sense of, and provide feedback on the presented pilot data. The GPs were a convenience, purposive sample based on two main principles. The first took into account the evidence surrounding single missed appointments. This describes missed appointments being more common in deprived, urban practices. The sample therefore included GPs who worked in deprived and affluent urban areas and a practice with a significant rural practice population from Scotland. Second, the sample included the views of frontline GPs and GPs who had a range of strategic roles in practice development and general practice management, locally and nationally. AEW and PW utilised their professional knowledge of GP networks and practice profiles to approach and recruit participants. Five out of twelve GPs contacted were able to attend the focus group. Each GP contacted reported that they felt this was an important topic worthy of attention. Barriers to attending were location of the focus group (conducted in Glasgow) and managing time away from other professional work. Additional file 1 describes each participant's characteristics. Detailed information about participants' practice characteristics was not collected. Three of the participants knew each other from their professional roles outside of clinical practice. AEW conducted the focus group and the analysis was conducted using Framework Analysis. Framework Analysis is a useful thematic analysis approach especially when considering a focussed topic like this one. Also in the context of being part of a larger mixed methods study, epistemologically its use was a good fit[22]. DAE attended the focus group and presented initial results from the 'proof of concept' pilot (described next) for discussion. Additional file 2 describes the topics covered in the focus group.

Proof of concept

Research that uses GP appointment data has not previously been conducted using the clinical recording systems in the Scottish NHS. A proof of concept pilot study was undertaken utilising the NHS Trusted Third Party (TTP) Albasoft with 67,705 patient records to determine whether retrieving appointment data was feasible, to refine other data parameters, and to inform the definition development as described in research question 1. An additional confidentiality control ensured that the research team did not know the identity of the recruited GP practices.

Additional file 3 describes the definition and role of TTPs.

Albasoft purposively recruited 10 Scottish practices on our behalf with the practice characteristics illustrated in figure 2.

Figure 2: Pilot practice recruitment

Data were cleaned and appointment rules applied to categorise appointments as attended or missed (DNA). Additional file 4 describes this process. This was primarily based on the 'in' and 'out' time recorded for each appointment. If this was recorded as '0' then the appointment was classified as Did Not Attend (DNA). For each patient the total number of appointments made during the three-year period was calculated as well as the number and percentage of appointments missed. Appointments that were recorded incorrectly in the system were removed at this stage. This included appointments where administrative records had remained open for over 24 hours, making it difficult to confirm that these were genuine appointments and not administrative errors. The pilot appointment rules are set out in table 1 below.

Data description	Reason for removal
total appointment time < 0 min	Record open for more than 24 hours

total waiting time < 0 min	Record open for more than 24 hours
appointment <2 min	Not a medical appointment
administrator slot	Not a medical appointment

Table 1 Rules to identify genuine appointments

Results

Focus group

Focus group participants reported making clear distinctions between patients who occasionally miss appointments and those who miss many. Patients who occasionally miss appointments do so because a crisis or another understandable circumstance has arisen; patients who serially miss appointments (SMA), described as missing more than two or three appointments can be easily identified by GPs. They were described as tending to have mental health, addiction, and/or social issues. They were described as high risk or vulnerable with concerns about their wider family. Patients who SMA were viewed as being different from the general GP population and being more likely to have 'chaotic' lifestyles associated with housing instability, money problems, a "*panicked lifestyle*"(P2). Patients who SMA were also described as being unable to manage GPs' expectation of them and fit into GPs' pre-determined slots. *"there's the occasional DNA which are quite normal and often those are quite significant [in total numbers for the practice] but the serial people I think that's a reflection of the chaos in their life whether that's you know- mental health or issues with the social functioning- and inability to manage our expectation of them- to fit into our pre-defined slots."* (P5)

All participants agreed with that view. However one participant also considered that not all patients who SMA can be viewed as high risk; that instead some patients do not value free health care. It was reported that some patients who SMA go on to book another appointment the next day; *"I don't think it's the value of*

the GP- I think it's the value of that appointment- I think the fact that it's, if you don't miss it, if you miss it is no big deal you just make another one" (P4).

Missed appointments were viewed as being more prevalent in practices in deprived settings, but occurred in affluent areas too. In the affluent setting they were important for a minority of marginalised, isolated patients with the same profile as described above-who were viewed as living 'chaotic' lives.

Practices represented in the focus group do not have protocols for managing patients who serially miss appointments (SMA) because response is dependent on the patient's context. GPs understood that SMAs usually mean patients with complex needs with workload implications for the practice. Strategies described were varied, including allowing patients only to book on the day; *"my impression is that deprived practices have a much higher percentage of on the day appointments because they skew it towards people that don't attend"* (P3), seeing the patient when they walk in, or the GP booking the follow up appointment for the patient- a relationship building strategy. This could still lead to patients missing an appointment, even just a couple of hours after it was made. It was reported that some practices do remove patients from their list for SMA and this created tension with other practices.

The focus group were also asked to comment on the results from the proof of concept initial data and they made recommendations about the full study design described in Figure 3.

Figure 3 Focus group recommendations for the full study design

Proof of concept

A pilot analysis of 67,705 patient records showed that while just over 60% of our sample missed no appointments, over 30% missed one or more appointment during the three-year period with nearly 10% of patients missing three or more appointments.

Assuming that our final sample provides a similar distribution, we will classify patients based on the number of appointments missed as follows:

***Never* missed appointments: 0 per year average over 3 year period**

***Low* missed appointments: <1 per year average over 3 year period**

***Medium* missed appointments: 1-2 per year average over 3 year period**

***High* missed appointments: >2 per year average over 3 year period**

Our sampling both in the pilot data stage and the final full study sample was conducted such that we were likely to get a representative sample of Scottish patients and practices. Because our pilot sample was large, it is appropriate to assume that this will scale-up accordingly for the full study. The distribution of missed appointments also suggested useful categories based on integer numbers of missed appointments per year. This will be helpful for policy and clinical stakeholders.

FULL STUDY PROTOCOL

Participants and study size

Our target recruitment of GP practices seeks to ensure that a spread of urban and rural practices, affluent and practices characterised by serving areas of blanket high socio-economic (Deep End) deprivation. The information request made to practices can be viewed in Figure 4.

Figure 4 Information request sent to target practices

Data sources and variables

GP data

The TTP has recruited the practices on the study team's behalf and will undertake some specific data aggregation before transferring the data securely to the National Safehaven for analysis. 'Safe Havens are specialised, secure environments supported by trained, specialist staff where data in electronic patient records can be processed and linked with other health data (and/or non-health-related data) and made available for analysis to facilitate research while protecting patient identity and privacy'[23]. These are: calculating urban rural classification, SIMD decile, categorising ethnicity into 'non BME (Black and Minority Ethnicity)', 'visibly BME', and 'non visible BME' and rounding travel distance to practice/emergency department for each patient record to the nearest kilometre. Once in a Safehaven, additional steps will be taken to ensure that acceptable anonymization principles are being applied, especially in relation to reporting of sensitive social vulnerability data and reporting of rare conditions.

A new data file containing the appointment history for each patient record will be generated, which will be merged with individual patient information (Additional file 4 describes this process based on our pilot data set).

Appointment validation and categorisation

Each appointment will be coded based on session type recorded by the practice (eg book on day appointments, or immunization clinic) and then further by professional type (eg GP partner, practice nurse). These descriptions are

determined by individual practices so categorisation will be conducted by the GPs in the research team. The appointment rules set out in the pilot study will be applied. A sensitivity analysis based on the time the appointment takes will then also be conducted by comparing a random sample of patient appointments as described in figure 5.

Figure 5 Random sample of GP appointments for validation and sensitivity analysis

The appointment rules will be refined based on this. The time interval cut-off for apparently attended appointments will be determined by utilising the time interval that most accurately matches to actual attended appointments. Slots designated non face to face appointments will then be removed leaving only attended and non- attended face- to- face appointments. The appointment categories described from the pilot study regarding non- attendance for all patients will then be applied to the yearly average number of missed appointments over the three year period to generate the four categories of patients for further analysis. Using an average over three years takes account of what is recognised in the frequent attenders (rather than non- attenders) literature- that patients' appointment behaviour may vary over time in relation to illness episodes or social crises[24].

Health and education data linkage

Linkage will be conducted as access permissions and data sets become available (figure 6). Each administrative data source is available for different time periods (e.g. hospital inpatients since 1981 and education outcomes since 2002) and this will be made explicit when interpreting the results. The TTP will provide the Safe Haven indexing team a file containing the GP dataset Community Health Index

(CHI) number and other patient identifiers. Every patient in the Scottish NHS has a CHI number, a unique identifier that is used as such across all NHS services in Scotland. This forms the cohort for the study. All data providers will supply identifiers to be probability matched to the study cohort by the Safehaven linkage team (based on CHI number and using other patient identifiers probabilistically for the small number of records where it is anticipated CHI will be missing), who will return a set of unique index numbers for those individuals successfully matched to the study cohort; each data provider will receive a different set of unique index numbers, and will use these index numbers as the basis of their data extract. Each data extract will be submitted to the Safehaven linkage team, who will replace the different index numbers with a common number across all files. This common number is the unique patient identifier that the research team will work from during analysis.

Figure 6 Proposed data sets for linkage with GP data

Bias

Accounting for patient turnover

This study seeks to ensure the inclusion of patients who are marginalised and who are often missing from health service studies. There is evidence of overlap between patients who miss appointments and those who are removed from practice lists[25], a recognition of the impact that geographical boundary areas have on patients who move around[26]; notwithstanding the gap in the literature about registration interruptions for patients who may go to prison or patients who remain unregistered once they are removed from GP practice lists. We will therefore summarise the numbers of patients joining and/or leaving their practice during the study period; with reasons where this information is available. We will seek to

provide a full analysis of the data available for these patients and compare these with the patients who are registered for the 3 year study period. Patients who are not registered with participating practices, and are being seen as 'temporary residents' by these practices, are excluded from the study. This is because these patients full clinical record is held by their registered GP so very limited information is available. Temporary residents tend to be people on holiday in the practice area but will include some people who would be considered marginalised.

Statistical methods

Our statistical analysis is based on the study being a retrospective cohort study. We will focus on two key areas; predictors of high rates of serial missed appointments, and serial missed appointments as a predictor of future patient outcomes.

Patient characteristics and practice characteristics may be associated with high rates of serial missed appointments. Analyses will initially be descriptive, summarising the rate of missed appointments in relation to the other factors recorded at the point of entry to the study. Associations with patient characteristics will be assessed as a whole, and in relation to different types of practices (e.g. separately in rural and urban practices). Subsequently, we will build regression models (Poisson or Negative Binomial), [27] to help understand how our categories of missed appointments are associated with patient and practice characteristics.

When considering other outcomes in relation to serial missed appointments, the missed appointment rate category (none, <1, 1-2, or >2 per year) will be the predictor variable. Appropriate regression models, according to the outcome, will be used to assess whether any associations with serial missed appointment rates

are independent of other patient- or practice-level factors. Conflicting interactions will be controlled for by using an 'offset term' in our negative binomial model which accounts for number of appointments made or any other relevant factors.

We also plan to measure relevant quantitative variables (described next) recorded *during* the study interval associated with having a lot of missed appointments. We will explore whether these *differ* from the predictive factors already recorded at entry to the study.

Quantitative variables

The following potential predictors of frequent non- attendance will be analysed:

Demographics

Patients' age, gender, minority ethnic group status (where available), deprivation decile, rural/urban split, number of address moves, distance lived from GP practice and distance from nearest A&E will be explored.

Health conditions

Health conditions will be reported using separate categories:

1. The incidence of multi-morbidity calculated from extracted Read codes based on previous counts in Scotland[28]

2. Descriptions of health conditions based on the priority 1 Read codes that GP practices in Scotland use to populate patients' key information summaries (KIS) for GP out of hours services. This is novel work as a coding structure has not previously been applied to these Read codes. Read codes are the clinical coding system used in UK general practice to record, clinical and administrative activity and diagnoses.

3. A count of psychotropic medicine prescriptions based on the British National Formulary. This is in order to describe levels of psychological morbidity that are not captured by diagnostic criteria.

4.These variables will then be compared to the ICD 10 coding data from patients' secondary care linked data compiled from hospital admissions and outpatient attendances. Diagnostic data from emergency department attendance was deemed not of sufficient quality to utilise.

Social Vulnerability

One aspect of this study which is particularly ground-breaking is our investigation of retrievable information about patients' social vulnerability. The Adverse Childhood Experiences (ACE) questionnaire[29] will be utilised as a template to match its nine descriptors of adversity to relevant Read codes in the patient's GP record. In addition, coding that maps the consequences of ACE will be analysed. A recent quantitative evaluation of Severe and Multiple Disadvantage will also be matched to GP Read codes. This examines the overlap of patients being homeless, in substance misuse services, or in prison over the preceding year[30]. Further, an exploration of additional Read codes that describe social vulnerability will be mapped. An anonymised text search linked to Read codes from the dataset will provide additional information about social vulnerability as it is recorded in the free text portion of GP records. Taken together, these will provide the first research evidence about the breadth and depth of social vulnerability recording by GPs.

Health care utilisation

Read coding in relation to cervical, breast and bowel screening attendance will be retrieved in addition to the proportion of patients who have had their blood

pressure checked and have participated in child health surveillance and vaccination programmes across the life course. A sub-analysis of utilisation of practice nurse and other health care professional's' appointments in the practice will also be conducted and include an exploration of the relationship between attending all primary care appointments and categories of non- attendance. This is because data from the GP focus group suggested there is overlap between patients who are serial non-attenders with patients who are very frequent attenders. We will therefore consider the rate of attending appointments as a potential predictor of the rate of non-attendance. Referrals that GPs make into other primary and secondary care services will also be analysed. Outpatient attendances, hospital admissions and utilisation of emergency departments, NHS 24 triage, GP out of hours, and ambulance services will also be analysed when linked data become available with a specific focus on how this relates to unmet need, for example how might GP appointment category relate to patterns of other health care utilisation between scheduled and unscheduled secondary care use.

Health care engagement

An analysis of GP Read codes and linked secondary care data will be carried out in the following categories:

1. Patients not attending primary and secondary care appointments

2. Patients refusing screening

3. Patients being exception-reported (ie excluded from the denominator population) from the Quality and Outcomes Framework (QOF) system for performance measurement in general practice

4. Practices' measures of non-engagement with care for long term conditions

5. Patients taking 'irregular discharge' from hospital (when patients leave against medical advice)

6. Patients not waiting to be seen in emergency departments

Family linkage

Diagnoses of children who are able to be linked through family linkage will be analysed based on their mother's appointment category. This is contingent on the child being included in the practice study population.

Education data

Attendance at school, exclusion from school, and educational attainment when leaving school will be made with approximately a sixth of our patient cohort for whom linked education data is available. This has the potential to inform future planning around earlier interventions to reduce serial missed appointments.

Practice level data

Each patient record will be allocated a unique practice ID enabling the research team to analyse each patient record output clustered by practice. This will be proportion of patients aged over 75, by ethnicity (proportion BME), patient rurality, patient level of deprivation decile, patient distance to practice, distance to A&E appointments offered/engaged, days from when appointment is made, multi-morbidity count, ACE score more than 4, Severe and Multiple Disadvantage score, hospital referrals, and proportion of each appointment category by practice. These analyses and output will be refined as the study proceeds taking patient level findings and multilevel modelling that characterises the respective contributions of practice- and individual-level factors to missed appointment patterns.

Health outcomes

Mortality data regarding date and cause of death will be utilised from GP and linked data. This will sit alongside additional linked obstetric outcomes (from the Scottish Birth Record) for relevant women.

Table 2 summarise the quantitative variables for analysis

Data categories	variables
Patient demographics	Age Sex ethnicity Count of address moves Distance to practice Distance to A&E SIMD decile Rural8 score
Health conditions	Multi-morbidity count Priority 1 read codes Psychotropic medication prescribing (BNF chapter) Secondary Health care diagnoses (inpatient and outpatient)
Social vulnerability	Adverse Childhood Experiences Severe and Multiple Disadvantage General social vulnerability coding frame
Health care utilisation	Breast screening Bowel screening Cervical screening BP checked Child health surveillance Vaccinations Practice nurse appointments Other health care professional appointments Primary care attendance distribution Hospital referrals Outpatient attendances Hospital admissions emergency departments attendance NHS 24 triage GP out of hours ambulance services callouts
Health care engagement	DNA codes Refused screening QOF exempt Inappropriate use codes Self-discharge codes
Study exit	Patient death Patient moved practice
Family linkage	Secondary health care linkage with mother and

	child
Education data	School attendance School exclusion School attainment
Health outcomes	Cause of death
GP Practice characteristics	Practice list size Patient age distribution Ethnicity category distribution Patient rur8 score distribution Patient SIMD score distribution Patient distance to practice distribution Patient distance to A&E distribution Number of appointments offered/patients engaged past 3 years distribution Number of days since appointments made distribution Patient multi-morbidity score distribution Patient ACE score distribution Patient SMD score distribution Patient hospital referrals distribution Primary care attendance pattern distribution

Table 2 Summary of quantitative categories and variables

ETHICS AND DISSEMINATION

This pathfinder linkage retrospective cohort study is necessarily complex in design and implementation because although cross-sectional it seeks to take a life course approach and follow the patients' journey through the health care system. Careful attention and significant resource has been devoted to the consideration of patient privacy and confidentiality. This has been integrated throughout the design of the study alongside the necessary data access and handling permissions. Additionally a study of this nature, which involves stakeholders across the NHS and other public services, requires a flexible time frame to allow access to raw data and to share findings between members of the research team based in several institutions. The proof of concept pilot did not require ethical approval because it was considered service evaluation with the agreement we would not publish any results about the practices who took part. Ethical permission to conduct the GP focus

group and publish the results was obtained by the MVLS ethics committee, University of Glasgow (ref 200140181). A letter of comfort was obtained from the West of Scotland NHS ethics committee and the MVLS ethics committee confirming that the full study did not need health service ethics permissions. Multi- site NHS R&D approval for the full study was obtained for all Scottish Health Boards (NRS16/186358).

Due to the sensitive nature of administrative data from the NHS and public education system in Scotland, the datasets generated and/or analysed during the current study will not be publicly available. They have been made available to the research team under controlled access and strictly for the purposes of this research study only. Summary data, at the level of disclosure checked output from the National Safehaven and statistical code, can be requested from the corresponding author on reasonable request.

Planned outputs

Alongside peer reviewed academic papers reporting the findings described above, the following additional outputs are planned.

Data Visualisation

Several web pages will be built to sit alongside key results. This will allow for the rapid construction of interactive data visualisations which will be created using "Shiny"[31], a web application framework for R which is the statistical software used for the study analysis. A simple platform will allow researchers and collaborators to interact with the analyses in real-time and generate custom tables and graphs as required. It can also provide non-experts with access to simple and complex statistical analysis using a point-and-click interface. This will not rely on raw data and will simply pull information from the summary descriptive analyses.

Case Studies

We also intend to use case studies to develop and illustrate our findings throughout the course of all our analyses. For example, we will be able to identify typical patient profiles of those who appear to miss many appointments in a very short period of time and compare these events with short and long-term health outcomes.

Conclusion

We shall identify key factors associated with serial missed appointments ranked in order of importance as described above, but given the large sample size we shall also be able to consider potential interactions that might predict serially missed appointments.

Finally, this approach also explores the theory that low engagement with health care should be viewed as a health harming behaviour, and will inform the debate about tackling health inequalities at the health service delivery level. Moving from theory into application, the results will allow us to better understand and develop future interventions to reduce serial missed appointments.

REFERENCES

1. Commission on Social Determinants of Health. Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health. Geneva: World Health Organisation, 2008.

2. George A, Rubin G. Non-attendance in general practice: a systematic review and its implications for access to primary health care. *Family Practice* 2003;20:178-84.

3. Campbell K, Millard A, McCartney G, et al. Who is least likely to attend? An analysis of outpatient appointment 'Did Not Attend' (DNA) data in Scotland. Edinburgh: NHS Health Scotland, 2015.

4. Neal RD, Hussain-Gambles M, Allgar VL, et al. Reasons for and consequences of missed appointments in general practice in the UK: questionnaire survey and prospective review of medical records. *BMC Family Practice* 2005;6(47)

5. Mitchell AJ, Selmes T. Why don't patients attend their appointments? Maintaining engagement with psychiatric services. *Advances in Psychiatric Treatment* 2007;13:423-34.

6. Ciechanowski PS, Russo J, Katon W, et al. Where is the patient? The association of psychosocial factors and missed primary care appointments in patients with diabetes. *General Hospital Psychiatry* 2006;28:9-16.

7. Lehmann T, Lehmann A, Olivet M, et al. Missed appointments at a Swiss university outpatient clinic. *Public Health* 2007;121:790-99.

8. Neilson KM, Faergeman O, Foldspang A, et al. Cardiac rehabilitation: health characteristics and socio-economic status among those who do not attend. *European Journal of Public Health* 2008;18:479-83.

9. Cashman S, Savageau J, Lemay C, et al. Patient health status and appointment keeping in an urban community health center. *Journal of health care for the Poor and Underserved* 2004;15:474-88.

10. Masuda Y, Kubo A, Kokase A, et al. Personal features and dropout from diabetic care. *Environment, Health Prevention Medicine* 2006;11:115-9.

11. Gucciardia E. A systemic review of attrition from diabetes education services: strategies to improve attrition and retention research. *Canadian Journal of Diabetes Care* 2008;32(5365)

12. Murdock A, Rogers C, Lindsay H, et al. Why do patients not keep their appointments? Prospective study in gastroenterology outpatient clinic. *Journal of Royal Society of Medicine* 2002;95:284-6.

13. Farley R, Wade T, Birchmore L. Factors influencing attendance at cardiac rehabilitation among coronary heart disease patients. *European Journal of Cardiovascular Nursing* 2003;2:205-12.

14. Woodward V. Non attendance at a nurse-run asthma clinic. *Journal Clinical Nurse* 1998;7:380-1.

15. Killaspy H, Banerjee S, King M, et al. Prospective controlled study of psychiatric out-patient non- attendance, Characteristics and outcome. *British Journal of Psychiatry* 2000;176:160-5.

16. Health and Social Care Information Centre England. Provisional monthly topic of interest: outpatient did not attend (DNA) appointment. London, 2014.

17. Watt G. Deep End Report 19. Access to Specialists. The 19th meeting of 'General Practitioners at the Deep End'. University of Glasgow, 2013.

18. Shonkoff JP, Garner AS, The Committee On Psychosocial Aspects Of Child And Family Health COEC, Adoption, And Dependent Care, And Section On Developmental And Behavioral Pediatrics,. The Lifelong effects of Early Childhood Adversity and Toxic Stress. *Pediatrics* 2012; 129.

19. Hussain-Gambles M, Neal RD, Dempsey O, et al. Missed appointments in primary care: questionnaire and focus group study of health professionals. *British Journal of General Practice* 2004;54:108-13.

20. Tudor Hart J. The Inverse Care Law. *The Lancet* 1971;297:405-12.

21. Kitzinger J. Qualitative Research: Introducing focus groups. *BMJ* 1995;311(7000):299-302.

22. Ritchie J, Lewis J. Qualitative Research Practice. A Guide for Social Science Students and Researchers. first ed. London: Sage Publications 2003.

23. The Scottish Government. Charter for Safe Havens in Scotland: Handling Unconsented Data from National Health Service Patient Records to Support Research and Statistics. Edinburgh2015 [Available from: http://www.gov.scot/Publications/2015/11/4783/2].

24. Smits FTM, Brouwer HJ, van Weert HCP, et al. Predictability of persistent frequent attendance: a historic 3-year cohort study. *British Journal of General Practice* 2009;59:114-19.

25. Williamson AE, Mullen K, Wilson P. Understanding “revolving door” patients in general practice: a qualitative study. *BMC Family Practice* 2014;15(1):1-11.

26. O'Reilly D, Steele K. The removal of patients who live outside the practice boundary. *British Journal of General Practice* 2005;55:384-86.

27. Zuur AF, Ieno EN. A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution* 2016;7(6):636-45.

28. Barnett K, Mercer SMN, Watt G, et al. Epidemiology of Multi-morbidity and implications for health care, research and medical education: a cross sectional study. *The Lancet* 2012;380((9836)):37-43.

29. Bellis MA, Ashton K, Hughes K, et al. Welsh Adverse Childhood Experiences (ACE) Study. Adverse Childhood Experiences and their impact on health-harming behaviours in the Welsh adult population. NHS Wales, 2015.

30. Bramley G, Fitzpatrick S, Edwards J, et al. Hard Edges Mapping Severe and Multiple Disadvantage England. England: Lankelly Chase Foundation, 2015.

31. Ellis DA, Meridian HL. Thinking outside the box: developing dynamic data visualizations for psychology with Shiny. *Frontiers in Psychology* 2015;6

AUTHORS' CONTRIBUTIONS

AEW is principal investigator for the study. DAE, PW and AMcC are co-investigators on the study and RMcQ is the research assistant. AEW conceived and developed the initial research proposal, reviewed the literature, conducted and analysed the pilot focus group, contributed to analysis and interpretation of the quantitative pilot data, developed the predictors, outcomes and associations of interest and led on writing the paper. DAE supported the development of the initial research proposal, reviewed the literature, conducted and analysed the quantitative pilot data, developed the statistical and output plan, and contributed to writing the paper. PW supported the development of the initial research proposal, reviewed the qualitative and quantitative pilot results, reviewed the statistical and output plan and contributed to writing the paper. RMcQ reviewed the statistical and output plan and contributed to writing the paper. AMcC provided expert statistical input to the study as it was developed, reviewed the statistical and output plan and contributed to writing the paper.

All authors read and approved the final manuscript.

FUNDING STATEMENT

This study is supported by a Scottish Government Chief Scientist Office research grant (CZH/4/41118) with Safehaven and data linkage costs supported in lieu by the DSLS at Scottish Government. These funding bodies had no role in the design of the study, or collection, analysis, and interpretation of data or in writing the manuscript.

COMPETING INTERESTS

The authors declare they have no competing interests.

ACKNOWLEDGEMENTS

We would like to acknowledge all GP practices and GPs who took part in the pilot study. Also colleagues at Scottish Government who are supportive of the study in a variety of ways especially Ellen Lynch in the Health Analytics Division. Dave Kelly's technical and procedural expertise, wisdom and patience as director of our TTP Albasoft Ltd, underpins all of what has been achieved to date.

FIGURE LEGENDS

Figure 1: Study research questions

Figure 2: Pilot practice recruitment

Figure 3 Focus group recommendations for the full study design

Figure 4 Information request sent to target practices

Figure 5 Random sample of GP appointments for validation and sensitivity analysis

Figure 6 Proposed data sets for linkage with GP data

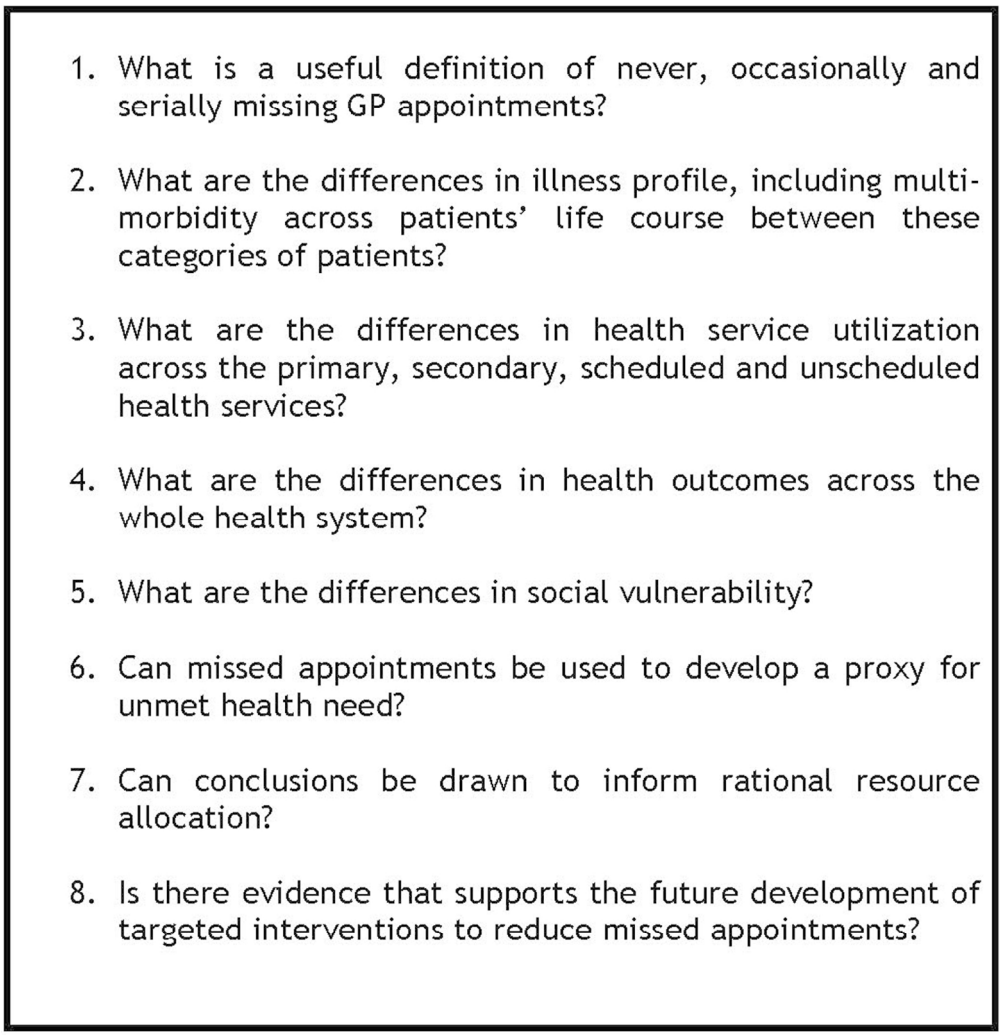


Figure 1: study research questions
Figure 1: study research quest
131x134mm (300 x 300 DPI)

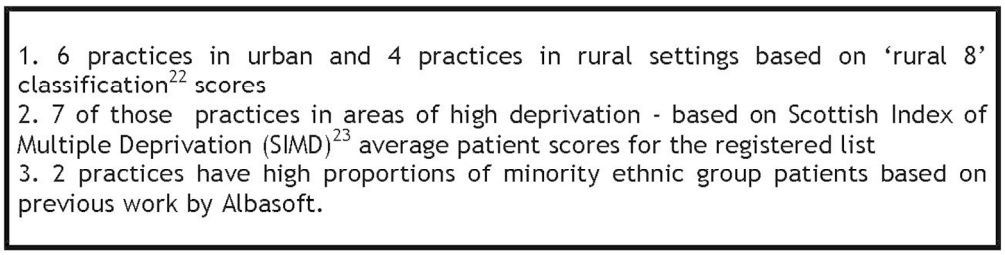


Figure 2: pilot practice recruitment
Figure 2: pilot practice recru
164x41mm (300 x 300 DPI)

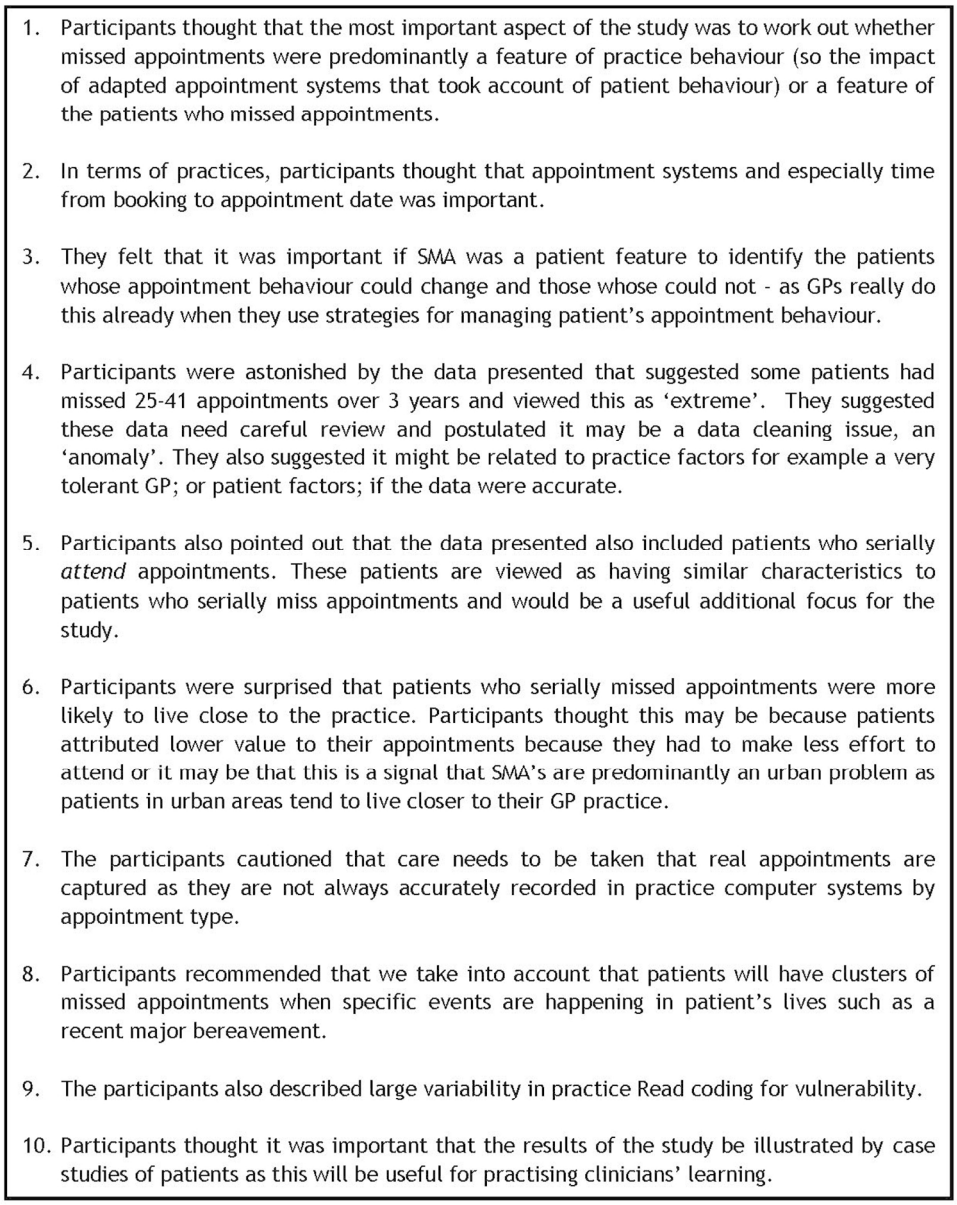


Figure 3: focus group recommendations
Figure 3: focus group recommen
157x197mm (300 x 300 DPI)

The data we are seeking are from the patient records of patients who have scheduled a GP appointment in the 3 years preceding the data extract date:

1. Dates of GP appointments

a. with missed or attended codes, b. session type, c. type of practitioner (e.g. GP, Practice Nurse), d. number of days since appointment made, e. unique practice ID

2. Patient demographic data

a. Age, b. Sex, c. Ethnicity, d. data zone, e. count of patient address moves , f. distance to practice (rounded to km), g. distance to hospital with an emergency department (rounded to km)

3. Clinical and prescribing data, Selected Read codes:

a. priority 1 coding (important health conditions that GPs code for export into the electronic care summary), b. long term condition diagnoses[2] (ref) c. patient vulnerability and adversity factors (ref), d. health screening (breast, bowel, cervical, BP, child health surveillance), e. exception coding (hospital referrals, DNA codes, refused screening, Quality and Outcomes (QOF, payment for targeted long term conditions management) exemption reporting, inappropriate use codes, self- discharge codes), f. specific prescribing information from BNF Central Nervous System (CNS) chapter and additional prescriptions specific to 40 long term diagnoses in b.)

4. Exit codes:

a. patient death, b. patient moved practice.

Figure 4: information request to target practices
Figure 4: information request
146x120mm (300 x 300 DPI)

Figure 5: random sample of appointments for sensitivity analysis
Figure 5: random sample of app
130x55mm (300 x 300 DPI)

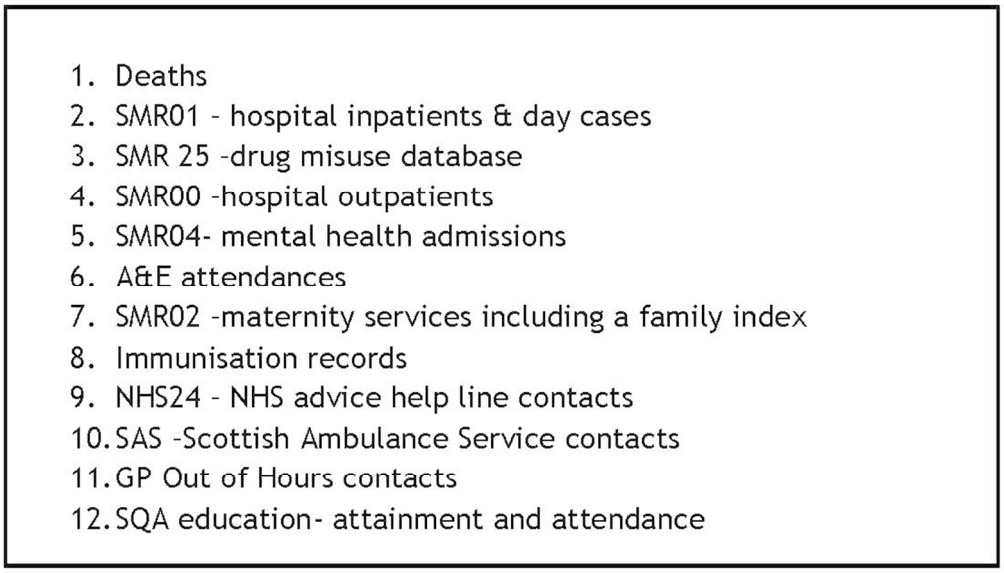


Figure 6: proposed linkage datasets
Figure 6: proposed linkage dat
140x79mm (300 x 300 DPI)

Additional file 1: GP focus group participant characteristics

Participant characteristics	Practice setting	Other work roles
• 4 male and 1 female GP • All aged 40-55 years old	• 3 high urban deprivation • 1 urban high affluence • 1 mixed semi-rural with pocket deprivation	• 1 clinical director of a Health and Social Care Partnership • 1 Local Medical Committee member • 1 clinical lead for a national innovation project • 2 with strategic Royal College of General Practitioner roles • 2 members of the 'GPs at the Deep End' steering group

Additional file 2: GP Focus Group Interview Schedule

Introductions:

Name, how long in clinical practice, time in your job, brief description of practice setting.

A priori knowledge/experience of missed appointments

Are missed appointments important? If so why? If not why not?

Can a distinction be made between patients who occasionally miss and those who serially miss GP appointments? If so what are those distinctions? Are they important?

How do you make that distinction in clinical practice? (probe distinctions between individuals and practice settings)

What does it mean for you, your practice and patients? Specifically patients who serially miss?

Present proof of concept provisional data (data cut offs, patient profiles)

What does this data tell us about the issue of serial missed appointments?

What are the obvious things it tells us? What are the surprises? Why?

Do you think it misses important aspects of what you think about the issue? Why might that be?

If we present these options about what a definition of a patient who serially missed appointments compared to one who occasionally does, which one do you think is most accurate? Why?

Is there more information that we should look for before deciding we have a definition? What should that be?

Conclusion

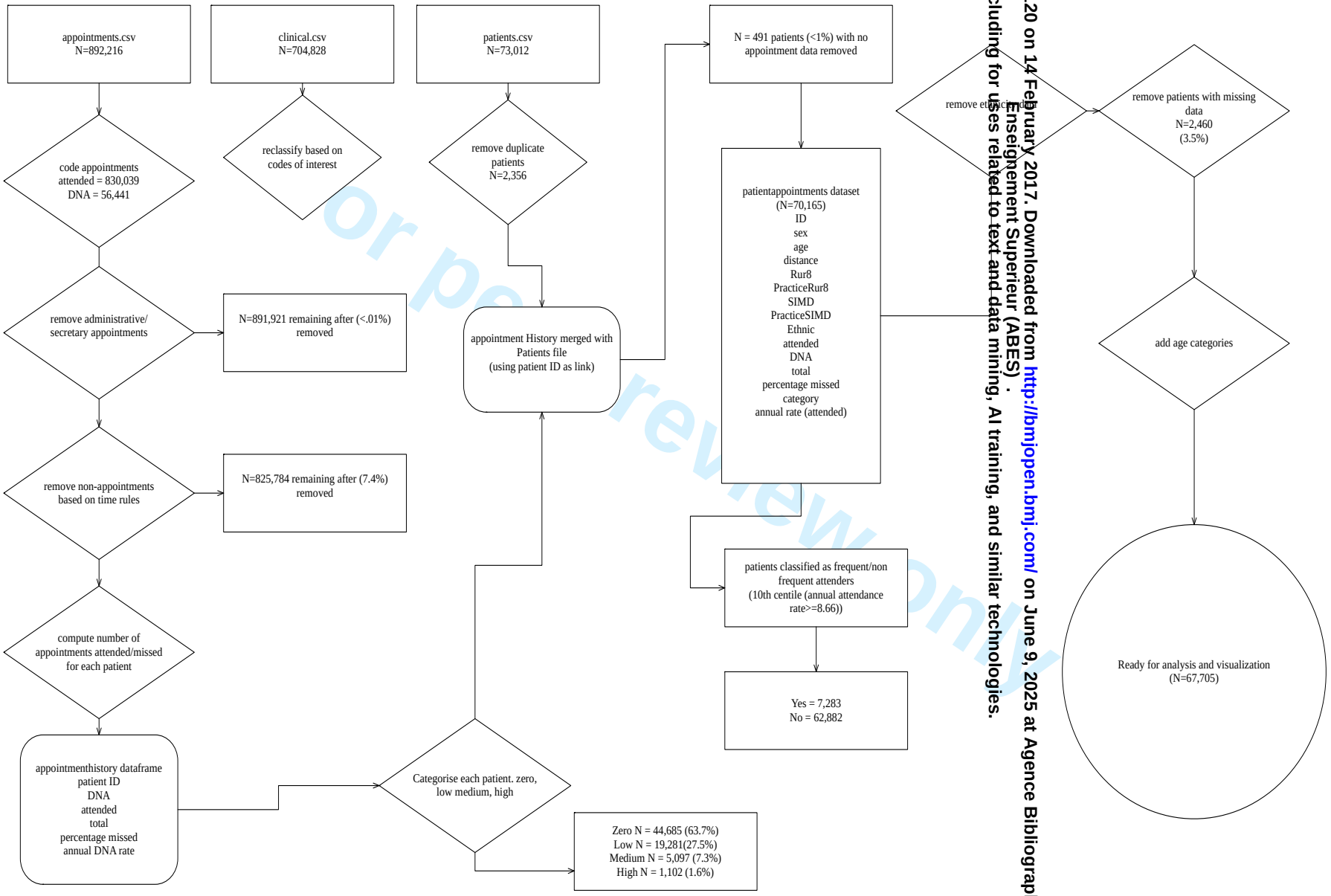
Finally, are there aspects of missed appointments and the definition development we have worked on today that we have not yet covered and you would like to tell us about?

Additional file 3: Definition and role of TTPs

With the increasing demand for statistical, research and service planning information from primary care records a solution is required to reduce the exposure of patient and clinician information to the requesting organisations to a minimum. The recommended (Information Commissioners Office) method of achieving this is by using a trusted third party (TTP) as an intermediary between organisations, which significantly reduces the number of individuals with access to identifiable information. In this case the TTP's role is to provide the technical skills to extract the required information from the Data Controllers electronic records and process this into a form that is both fit for purpose and complies with principal 3 of the data protection act. This may require the removal/replacement of identifiers (anonymisation /pseudo-anonymisation) or the use of redaction techniques when only statistical information is required prior to release of information to the beneficiary.

A TTP is required to operate to strict guidelines as it may only processes data in accordance with instructions from the data controller and to a specification previously agreed by both data requestor and data provider. The TTP acts as a Data Processor on behalf of the Data Controller and abides by the principles defined in the data protection act. It is registered as a data processor with the ICO, provides a secure storage facility which operates procedural, physical and electronic access controls to protect the data it processes and has no specific interest in, not is affiliated to any organisation that has an interest in any data provided. Albasoft maintain a secure data processing and storage facility at the Centre for Health Science adjacent to Raigmore hospital in Inverness, this facility is solely hosted on the NHS network. No information is transferred out with the

NHS network. Its existing middleware platform Escro is an advanced practice based reporting system and is used to securely process data locally at the practice before transferring the results to their secure repository. Albasoft has an established track record as a TTP for the Scottish Therapeutics Utility and increasingly in supporting NHS research. In our study, Albasoft have established data sharing agreements with Scottish GP practices for computerised access to the GP practice data.



STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2-3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-6
Objectives	3	State specific objectives, including any pre-specified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	Pilot 8-10 Full study 14
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	n/a
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	15-17
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*15-17*
Bias	9	Describe any efforts to address potential sources of bias	18
Study size	10	Explain how the study size was arrived at	
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	14
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	19-20
		(*b*) Describe any methods used to examine subgroups and interactions	18, 19-20
		(*c*) Explain how missing data were addressed	16-17
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	n/a

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy 19-20	
		(*e*) Describe any sensitivity analyses	16
Results			N/A
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	-
		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	-
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	-
		(b) Indicate number of participants with missing data for each variable of interest	-
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	-
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	-
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	-
		Cross-sectional study—Report numbers of outcome events or summary measures	-
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	-
		(*b*) Report category boundaries when continuous variables were categorized	-
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	-
Discussion			-
Key results	18	Summarise key results with reference to study objectives	-
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	-
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	-
Generalisability	21	Discuss the generalisability (external validity) of the study results	-
Other information			-
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	31-32

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.