Full title of project

Accounting for multimorbidity, competing risk and direct treatment disutility in risk prediction tools and model-based cost-effectiveness analysis for the primary prevention of cardiovascular disease and major osteoporotic fracture

Bruce Guthrie, Professor of Primary Care Medicine, University of Dundee

Dan Morales, Discovery Fellow, University of Dundee

Alex Thompson, Research Associate, University of Manchester

Peter Donnan, Professor of Epidemiology and Biostatistics, University of Dundee

Katherine Payne, Professor of Health Economics, University of Manchester

Sarah Davis, Senior Lecturer in Health Economics, University of Sheffield

Chief Investigator: Bruce Guthrie, Population Health Sciences Division, University of Dundee, Mackenzie Building, Kirsty Semple Way, Dundee DD2 4BF. b.guthrie@dundee.ac.uk

Summary of Research

Background: Medicines for the primary prevention of disease are taken long-term and often for life, which makes decisions about starting such medicines very important to patients and the NHS. Clinical guidelines increasingly make risk-stratified recommendations, where treatment offers vary by predicted baseline risk of the outcome to be prevented (for example, cardiovascular disease (CVD) or osteoporotic fracture). Treatment thresholds recommended by guidelines are usually informed by model-based cost-effectiveness analysis (CEA), and implementing risk-stratified recommendations requires clinicians and patients to use risk prediction models. These models are therefore of much more than academic interest since they are increasingly central to both guideline development and clinical decision making. However, existing models are potentially problematic in older people and those with life-limiting multimorbidity. Risk prediction models do not usually account for competing risks of death from other causes. This means that they are likely to over-predict CVD or fracture risk in people with multimorbidity, who will often die of other causes. Existing cost-effectiveness models do account for average competing risk but average population cost-effectiveness may not apply to people with multimorbidity in whom competing mortality risks are much higher, and additionally the relatively small lifetime QALY gains from primary prevention may be offset by unaccounted for disutility relating to treatment burden.

Overall aim: To improve the evidence generated from risk prediction models and model-based cost-effectiveness analyses to better inform decision-making for selecting primary prevention treatments in two selected case studies in people with multimorbidity.

Methods: The study has three elements. (i) Derivation of new models for predicting the risk of incident CVD and osteoporotic fracture using Clinical Practice Research Datalink (CPRD) data to model both the targeted outcomes (incident CVD and osteoporotic fracture) and unrelated death (the competing risk), with internal cross-validation in CPRD to examine discrimination and calibration. We will compare the new model with the most widely used existing CVD and fracture risk prediction models (QRisk2 and QFracture) neither of which accounts for competing risks. ii) Elicitation of Direct Treatment Disutilities (DTD) for the general population and for people treated with statins or bisphosphonates, using time-trade off and best-worst scaling, with design informed using systematic literature reviews and semi-structured interviews with 15 patients and clinicians. (iii) Adaptation of the model-based CEA for lipid modification and bisphosphates used by NICE to quantify the impact of accounting for competing risks and/or DTD on relative cost-effectiveness and effectiveness measured by lifetime QALY gain.

Expected outputs and impact: Better risk prediction in people with multimorbidity by accounting for competing risks will improve the targeting of preventative medicines to those who are most likely to benefit, and potentially reduce the use of low-benefit but potentially harmful medicines. Accounting for variation in effectiveness and cost-effectiveness by appropriate subgroup analysis in model-based CEA to account for competing risks and DTD will ensure that guideline recommendations for preventative medicines are more appropriately nuanced to the important population of people with multimorbidity. We will ensure that the findings are appropriately disseminated to the full range of interested clinical, guideline developer, public and academic audiences. As well as academic outputs (the NIHR final report, conference presentations, peer-reviewed papers), we will produce summaries for clinical and guideline development audiences, and offer workshops/presentations to NICE and SIGN with whom we have established relationships from our previous work. We will make our algorithms and methods publicly available in suitable forms to facilitate external validation and implementation in clinical IT systems, as well as provide a web-based calculator with guidance on appropriate interpretation.

Background and Rationale

Clinical guidelines have a key role in defining and disseminating best practice, but have been criticised for their focus on single conditions which is potentially problematic in the majority of people

with long-term conditions who are multimorbid.[1-6] A particular focus of this critique is in the use of medicines for prevention of future disease, where treatment is usually long-term and often lifelong. Many preventative medicines have relatively small mean expected benefit to individuals especially when used for primary prevention, and ensuring that treatment is appropriately targeted to maximise net benefit is therefore critical. Since relative risk reductions from treatment are generally assumed to be constant across all populations, a major determinant of an individual's expected benefit is their baseline risk of the outcome being prevented. Other things being equal, the higher an individual's baseline risk then the greater the benefit they can expect, and guidelines increasingly make risk-stratified treatment recommendations. Model-based cost-effectiveness analyses used to inform guideline recommendations about preventative treatments will often examine cost-effectiveness for different levels of baseline risk to help guideline development groups decide on appropriate baseline risk thresholds at which preventative medicines should be recommended. Risk prediction and model based cost-effectiveness analyses are therefore increasingly used to underpin the development and implementation of recommendations for preventative medicines. Although both types of model are highly technical, they are therefore far from only being of academic interest since they directly influence decisions about the lifelong treatment of very large numbers of people.

Risk prediction models

Based on a model-based cost-effectiveness analysis, the recently updated NICE lipid modification guidelines recommend clinicians use the QRisk2 risk prediction tool to estimate 10-year cardiovascular disease (CVD) risk and offer statins to all people for whom predicted risk over that period is >10%.[7] Since age is the strongest determinant of CVD risk virtually all people in the UK cross the QRisk2 estimated 10% threshold by age 65 years irrespective of classic CVD risk factors. The guideline therefore effectively recommends that all older people without cardiovascular disease should be offered a statin, implicitly recommending medicating all older people for the rest of their lives. Similarly, NICE recommends that clinicians use fracture risk prediction tools such as QFracture to help identify patients at high risk of osteoporotic fracture to inform decisions about further investigation with bone mineral density measurement or starting preventative treatment.[8] In the case of fracture, NICE do not currently recommend a risk threshold at which to start treatment, although other UK national guidelines recommend treatment thresholds based on risk of osteoporotic fracture,[9] and there is a NICE technology appraisal of bisphosphonates in progress[10] which is informed by a model-based cost-effectiveness analysis and which may recommend treatment thresholds based on predicted risk of fracture.

Risk prediction models are known to systematically over-predict baseline risk in people with life-limiting comorbidity and in older people with multimorbidity because they often have high (competing) risk of death from unrelated causes. Competing risks occur when an individual is at risk of death from an unrelated condition to the one being studied (for example, death from respiratory disease in a study of CVD or osteoporosis). Competing risk is a well-recognised problem in survival analysis and there are a number of established methods for dealing with it,[11-15] although these have only recently been applied to risk prediction model development.[16-19] In essence, conventional survival analysis and risk prediction tool development treats death from an unrelated condition as a right-censored event similar to loss to follow-up. Such prediction models therefore assume that censored individuals have the same risk of the outcome being examined as those remaining in the study, which clearly is not true for individuals who have died of unrelated causes. The estimates from conventional survival analysis are therefore only accurate in the notional world where people cannot die from other conditions, and overestimate risk of the outcome in those with competing risks.[11,13,14] In the context of 10-year prediction of the risk of first CVD event, the assumptions made by such models are reasonable for middle-aged people without a life-limiting condition, but are problematic in the presence of life-limiting comorbidity at any age or of multimorbidity in older people. Studies which have compared conventional and competing risk adjusted models have shown that conventional models on average over-predict CVD risk,[16,18-20] but have not explored the implications of over- (and

potentially under-) prediction in people with non-CVD morbidities. To our knowledge, the effect of competing risks has not been systematically examined in the context of fracture risk prediction, but we would expect them to be greater because fractures cause a much smaller proportion of total mortality than CVD.

There are important clinical implications of over-prediction of baseline risk because the patients concerned are both less likely to benefit than expected, and are likely to have comorbidity and co-prescribing which increases their risk of adverse drug events.[21] For example, although statins are remarkably safe drugs with few side effects experienced in clinical trials in mostly middle aged people, the only published trial of stopping statins in people with reduced life expectancy from a variety of causes found that cessation led to improved quality of life with no significant increase in cardiovascular events, consistent with net harm in the presence of competing mortality risks.[22] The first objective of this study is therefore to derive and internally validate new incident CVD and osteoporotic fracture risk prediction models for people with multimorbidity accounting for competing risks of death and compare performance with the QRisk2[23] and QFracture[24] risk prediction models (the most commonly recommended and used prediction tools for CVD and fracture in the UK).

Model-based cost-effectiveness analysis

Model-based cost-effectiveness analyses have become an integrated source of evidence in the development of clinical guidelines by decision-making bodies such as NICE. There is a growing recognition from NICE and other guideline developers and by users that the current single-disease approach for creating guidelines is problematic.[25,26] Since most people with any long-term condition and most people aged 65 years and over have multimorbidity,[1] there is a demand for guidance produced by NICE which accommodates for the complex needs of these populations.[6] Evidence based recommendations in national clinical guidance can only be useful when the evidence relates to the populations that is seen in consultations day-to-day. Failing to account for people with multimorbidity means that guidelines will be less useful, and risks that recommendations (rightly or wrongly) will not be adhered to. There is emerging evidence that taking account of multimorbidity will affect the total net benefit estimated for interventions compared with current practice, which in turn will influence whether or not to recommend the intervention for use in clinical practice. This evidence has indicated the need to account for important sub-groups of the total patient population and the role of heterogeneity in the evidence base for different types of patient.[27] Further evidence from our recently completed NIHR HS&DR funded Better Guidelines project has shown that accounting for plausible levels of competing mortality risks associated with multimorbidity and of direct treatment disutility (the impact on health status of taking a long-term preventative medicine irrespective of specific adverse effects of the actual drug) could also influence if, and when, overall net benefit is achieved.[28] Consequently, the methods used to structure and populate model-based cost-effectiveness analysis need to be re-considered in order to meet this new demand for evidence that takes account of multimorbidity.

Models used in cost-effectiveness analysis do account for competing risks in that they typically include both population age-specific total mortality and related cause-specific mortality in the model. However, treatments which are cost-effective on average may be less cost-effective (or judged not cost-effective) in important subgroups such as in people with life-limiting conditions in whom population average mortality underestimates true mortality and therefore likely overestimates treatment benefit.[27,29] In our recently completed NIHR HS&DR funded ‘Better Guidelines’ project, we explored the impact of varying competing risks of death using the NICE lipid modification cost-effectiveness model and found that plausible increases in competing risks significantly affected expected lifetime QALY gain and therefore potentially cost-effectiveness at different baseline risk thresholds for initiating statin treatment.[28] However, we could not find any published data on the distribution of competing risks across the population which meant that it was not possible to properly explore the implications of this for cost-effectiveness. Of note is that people with greater risk of CVD or fracture will typically also have greater competing risk of death, for example because smoking

causes both CVD and fatal respiratory disease of various kinds, and increasing age and nursing home residence is associated with both fracture risk and risk of death from many causes.

A second way in which cost-effectiveness models may be misleading is because they do not include all harms. Again, in the 'Better Guidelines' project we examined the impact of accounting for one type of harm which is currently ignored by existing models – direct treatment disutility (DTD). We define DTD as being the collective set of individuals' strength of preference for not taking a medicine long-term, which may arise for a number of reasons. Patients are likely to value negatively the inconvenience of obtaining prescriptions from GPs and collecting medicines from pharmacies, of taking medicines regularly, of having to attend for monitoring of various kinds, and of needing to modify their lifestyle to take the medicines, as is the case for bisphosphonates for example. For some patients, taking a regular medicine or other intervention for life is an unpalatable prospect in its own right in addition to the specific hassles of being on treatment. The concept of DTD is over and above the dis-benefit (harm) captured by attaching a disutility (negative impact on health status) associated with adverse drug events (ADEs) or the financial (out of pocket) costs for the patient. The disutility of ADEs is usually included in decision-models to some extent. However, the negative impact of taking a medicine long-term or for life, especially a preventative medicine with no obvious immediate benefits, is currently ignored in model-based cost-effectiveness analysis.

In the Better Guidelines project, we found that even very low plausible levels of DTD could significantly reduce or even reverse expected lifetime QALY gain in the context of statin treatment at the new NICE treatment threshold of 10% CVD risk at 10 years, where treatment had a slowly accruing and relatively modest lifetime net benefit.[28] There is a small published literature in this field, with the cost-effectiveness of several primary preventative treatments having been shown to be sensitive to even small levels of DTD or treatment burden,[30-32] but there is a need to better quantify DTD because DTD values have only been elicited in a small number of studies and there is uncertainty as to their magnitude and distribution.[33,34] DTD may also vary by treatment. For example, statins to prevent CVD have to be taken daily compared with weekly bisphosphonates to prevent fracture, but the routine for taking bisphosphonates is much more complicated (taken on an empty stomach with a significant quantity of water and with a requirement to stay upright for at least 30 minutes and not eat or drink for 30-60 minutes after ingestion).

For objectives 2 and 3, this study will therefore also elicit DTD values for taking lifelong statins and bisphosphonates from both general population and treated patient samples, and use these and data from the competing risk adjusted prediction models to examine how expected lifetime QALY gain and cost-effectiveness varies in the presence of different levels of DTD and competing risk.

Why this research is needed now

Examining the impact of competing risks in risk prediction and economic models, and better accounting for treatment harm from long-term preventative medicines are the key objectives of this project. Economic models are central to informing the decisions made by NICE guideline development groups about which treatments to recommend and when to recommend them, with a continued increase in the use of risk-stratified recommendations for preventative treatments. Risk-stratified recommendations require valid risk prediction models for implementation in decision-making for individual patients. Risk prediction and economic models are therefore of central importance in both guideline development and clinical decision-making for preventative treatments, but are problematic in people with multimorbidity in the face of competing risks, and often do not fully account for variable preferences for taking long-term treatment.[27,29] Given the scale of use of drugs for primary prevention, health need and sustained importance to the NHS are high, and many clinicians and patients have expressed concern about guideline recommendations which are often perceived as being over-encompassing, notably in relation to much wider use of statins. The study will therefore improve targeting of preventative treatment in people with multimorbidity, which is of major concern in

the face of rising rates of multimorbidity due to population aging and better survival from acute events such as heart attack, stroke and cancer.

Aims and objectives

The overall aim is to improve the evidence generated from risk prediction models and model-based cost-effectiveness analyses to better inform decision-making for selecting primary prevention treatments in two selected case studies in people with multimorbidity.

Objective 1: To derive and internally validate new incident CVD and osteoporotic fracture risk prediction models for people with multimorbidity accounting for competing risks of death, and compare performance with existing risk prediction models.

Objective 2: To quantify the magnitude, variation and distribution of Direct Treatment Disutility (DTD – the disutility incurred by taking a regular, long-term treatment irrespective of drug-specific side effects) in the general and statin or bisphosphonate treated populations.

Objective 3: To examine the effect of accounting for competing risks and Direct Treatment Disutility on clinical effectiveness and relative cost-effectiveness in the context of the use of statins and bisphosphonates for the primary prevention of CVD and osteoporotic fracture respectively.

The study will have three work-streams to match these objectives. Objective 1 will be delivered by a statistical modelling study carried out using the Clinical Practice Research Datalink, which is the longest established and single largest GP electronic medical records research dataset. A stated preference elicitation study of the general population and of treated patients will be used to quantify the size, variation and distribution of DTDs for objective 2. Objective 3 will be delivered by using data from the first two work-streams to adapt the model-based cost effectiveness analyses developed to inform the NICE lipid modification guideline published in 2014 and the NICE bisphosphonate technology appraisal which will publish in November 2015.

Research Plan / Methods

Design and theoretical/conceptual framework

The overall design is two linked statistical and economic modelling studies, combined with primary data collection to elicit direct treatment disutility values for use in the economic modelling. We therefore plan three linked sub-studies involving:

1) Statistical modelling study using Clinical Practice Research Datalink (CPRD) data to develop two new risk prediction models accounting for competing risks in people with multimorbidity;
2) Mixed methods study to elicit direct treatment disutilities (DTD);
3) Model-based cost effectiveness analysis drawing on the findings of the first two studies.

Methods for objective 1: risk prediction modelling

Objective 1: To derive and internally validate new incident CVD and osteoporotic fracture risk prediction models for people with multimorbidity accounting for competing risks of death, and compare performance with existing risk prediction models.

Rationale: Guidelines which make recommendations for the use of treatments for primary prevention increasingly make risk-stratified recommendations where treatment is only recommended if individuals have a predicted risk of the outcome being avoided which is above a particular threshold. A recent (and somewhat contentious) example of this is the new NICE lipid modification guideline which recommends offering statin treatment to everyone with a QRisk2 predicted 10-year risk of CVD of 10% or more, compared to previous guidance to offer statins at the 20% threshold. Risk-stratified recommendations are useful because they focus treatment on those who are most likely to benefit,

but may be problematic in older people and people with multimorbidity because existing risk-prediction models including QRisk2 and QFracture ignore competing risk of death from unrelated causes. This means they will overestimate CVD and fracture risk, particularly in older people and those with multimorbidity, but also in younger people with life-limiting comorbidity. However, it is unclear the extent to which competing risks would alter treatment recommendations, and which subgroups of the population being recommended for treatment are most affected, but both of these questions can be answered by analysing routine population-level data.

Design. Statistical modelling study to create and internally validate prediction models for incident fatal and non-fatal cardiovascular (CV) events and incident hip fracture, to compare their overall performance with risk prediction tools in widespread use in UK clinical practice, and to explore performance in people with different patterns of multimorbidity and consequential reduced life expectancy. Analysis will be carried out in SAS or Stata depending on the experience and skills of the employed researcher.

Setting/context. Statistical modelling will be carried out using data from the Clinical Practice Research Datalink (CPRD) which includes GP electronic medical records (EMRs) containing demography, diagnoses recorded using Read Codes (a hierarchical thesaurus of coded clinical terms used in UK primary care) and a number of clinical and laboratory values such as blood pressure and cholesterol. From 1997, CPRD GP data are linked to hospital admissions recorded in Hospital Episode Statistics (HES) and to deaths recorded by the Office for National Statistics (ONS) database. Currently, linked data is available for ~5 million patients and this is increasing as more practices consent to linkage. HES and ONS diagnoses are recorded using the International Classification of Disease coding system. Linkage to HES and ONS data is required to prevent the underestimation of events which has been shown to occur for up to 25% of acute myocardial infarction cases using only one source.[35] General practices and patients within CPRD meet defined quality standards to contribute data, and diagnoses within CPRD have high validity.[36,37] For example, the validity of acute myocardial infarction recorded in primary care medical records is 92%, whilst the validity of hip fractures and vertebral fractures is 91% and 88% respectively.[36,37] CPRD is an appropriate data source in which to conduct this study firstly because incidence rates recorded in CPRD are very similar to those in the QResearch population used to derive the QRisk2 and QFracture algorithms and secondly because these algorithms have been validated in CPRD data with almost identical performance as in the original QResearch data.[38] Approval to conduct the study will be sought from the Independent Scientific Advisory Committee (ISAC) for Medicines and Healthcare products Regulatory Agency database research.

Study population. For each of the outcomes (CV events and hip fracture), we will define an open cohort of permanently registered patients. Cohort entry will be defined as the later of one year after the patient enters CPRD or the 1st January 1998 or the patient's 30th birthday. For the cardiovascular disease model, people with CVD or taking a statin prior to cohort entry or with undated CVD codes or undated statin prescriptions will be excluded. For the hip fracture model, people taking a bisphosphonate prior to cohort entry or with undated bisphosphonate prescriptions will be excluded.

Outcomes. Incident CVD will be defined as the earliest of a GP-recorded Read Code for myocardial infarction, angina, other coronary or ischaemic heart disease, stroke or transient ischaemic attack (TIA) *or* HES-recorded ICD10 code for hospital admission with ischaemic heart disease or ischaemic stroke/TIA *or* death from CVD. Follow-up will otherwise be censored at the date of first statin prescription, date of deregistration, date of death or end of follow-up for the entire study. Incident hip fractures will be defined as the earliest of a GP-recorded Read Code for hip fracture *or* HES-recorded ICD10 code for hospital admission with hip fracture *or* death from hip fracture. Follow-up will otherwise be censored at the data of first bisphosphonate prescription, date of deregistration, date of death or end of follow-up for the entire study.

Risk factors predicting outcomes. For both outcomes, we will include variables already used in QRisk2 and QFracture since we wish to focus on the impact of competing risks rather than deriving completely new risk prediction tools (additionally both QRisk2 and QFracture already contain all variables known to be strongly and independently associated with outcome including age, sex, deprivation and ethnicity, as well as a range of other variables based on observed associations in similar GP datasets).[23,24]

Imputation for missing risk factors. Where appropriate, we will assume that non-recorded means that the patient does not have the condition of interest (for example, diabetes). However, some risk factor variables are known to have significant proportions of patients with missing data where such an assumption is not viable (for example, it would be inappropriate to assume that missing data for alcohol consumption means the patient is teetotal). For these risk factors, we will use multiple imputation to address this, assuming that data is missing at random (MAR), using either proc mi and proc mianalyse in SAS or multiple imputation using chained equations (MICE) in Stata. The number of imputations will be increased from the default of five where the proportion of missing reduces the relative efficiency to less than 80%. We will additionally examine varying the index date at which CVD or fracture risk assessment is done in order to minimise missingness while avoiding reducing available follow-up time excessively.

Deriving competing risk regression models. The common practice of developing the model on only a random fraction of the data available and then internally validating it on the remaining fraction is inefficient as it wastes information. Bootstrapping will be used where samples are drawn with replacement and a series of models are fitted on the full sample size. Bootstrapping has been shown to better account for all sources of model uncertainty including variable selection as well as resulting stable regression coefficient estimates.[39] Using the multiply imputed dataset, we will use Fine and Gray competing risk adjusted Cox proportional hazard models to estimate the associations between risk factors and outcomes, accounting for competing risk of death (non-CV death or non-hip fracture death), and therefore use the cumulative incidence estimator for CVD and hip fracture outcomes rather than the Kaplan-Meier estimator which is biased in the presence of competing mortality risks.[13,14] Separate models will be developed for men and women and we will use the Akaike Information Criteria (AIC) to evaluate model fit since this penalises models with more variables, with the additional advantage that it is equivalent to cross-validation. Separate models will be developed for men and women. Continuous variables will be centred and non-linear risk relations accounted for using fractional polynomials if required. The assumption of proportional hazards will be checked. Regression coefficients from the final model will be used as weights to predict CVD and hip fracture outcomes in individuals by combining them with the baseline survivor function to estimate CVD and hip fracture at annual increments up to 10 years.

Internally validating competing risk regression models. Internal validation will be achieved through the efficient process of bootstrapping. We will compare predicted CVD event rates with observed rates estimated using competing risk life tables, and calculate the D statistic, R^2 and area under the receiver operating curve or c-statistic. We will examine calibration by comparing mean predicted risk in equal tenths (deciles) and equal twentieths (vigintiles) of the population with mean actual risk. We will additionally examine sensitivity and specificity of the risk score at current (10%) and previous (20%) 10-year estimated risk thresholds.

Comparison with existing risk prediction models. In parallel with the development of the competing risk adjusted risk prediction models, we will implement QRisk2 and QFracture in CPRD using the published algorithms. To minimise the risk for error, we will purchase a commercial licence for each which includes the look-up tables (eg for deprivation score) used in the published online tool and risk factor weights based on the most recent dataset. We will then examine discrimination and calibration of QRisk2 and QFracture in CPRD using both observed CVD and hip fracture events from conventional life tables and observed rates using competing risk life tables.

We will compare the competing risk adjusted models and QRisk2 and QFracture by first examining the extent to which risk estimates differ between tools in individuals, in terms of the difference in estimated absolute risk at 10 years for different deciles and vigintiles of predicted risk. We will then examine reclassification, which is the extent to which different models classify different patients as high, intermediate or low risk, since this provides a better indication of the extent to which accounting for competing risks leads to changes in which individual patients are recommended for treatment. For CVD, we will examine reclassification using <10%, 10-20% and >20% 10-year CVD risk, since 10% and 20% are the new and old NICE recommended thresholds for initiating statin treatment respectively. For hip fracture, there is no NICE guideline, and we will define thresholds based on the final version of the draft NICE bisphosphonate technology appraisal if they recommend treatment thresholds and National Osteoporosis Guideline Group guidelines if not.[9,40] Both of these approaches will be initially taken in the entire population, but we will then further examine which patients have large changes in estimated risk after accounting for competing risk, and which patients change from being recommended treatment at different thresholds to not being recommended treatment, in terms of condition count (which we will implement using the same approach as our previous multimorbidity work[1]), drug count (as a more easily routinely measured proxy for multimorbidity, both total drug count and 'other condition' drug count ie excluding CV and osteoporosis drugs as appropriate[2]) and frailty (using the electronic frailty index, which has recently been developed and externally validated using UK general practice electronic data[41]). Finally, we will further examine the impact of accounting for competing risks in selected groups of people with high competing risks of mortality due to other conditions, such as people with COPD, people with cirrhosis and chronic liver disease, people resident in nursing homes, and people aged 80 years and over.

Methods for objective 2: elicitation of values for direct treatment disutility

Objective 2: To quantify the magnitude, variation and distribution of Direct Treatment Disutility (DTD – the disutility incurred by taking a regular, long-term treatment irrespective of drug-specific side effects) in the general and statin or bisphosphonate treated populations.

Rationale: There is a growing evidence-base that taking a specific treatment, particularly one requiring long-term use for a chronic condition, can cause inconvenience or "disutility" to a patient which is exclusive to the unwanted harms, adverse outcomes or specific effects of the treatment.[30,32,33,42-46] Direct treatment disutility (DTD) may have particular relevance for long-term medication use, such as statins for the primary prevention of cardiovascular disease (CVD) and bisphosphonates for osteoporosis since the benefits of treatment are typically small and accrue over long periods. Direct treatment disutility can be defined as representing an individuals' strength of preference for *not* taking the medicine.[47] Current model-based cost-effectiveness analyses ignore the impact on health status of the act of taking a tablet for chronic preventative treatments, and so assume that these cause no inconvenience or harm beyond specific adverse effects or out of pocket costs, or in other words, that such treatments have no direct treatment disutility (DTD).[47] Existing empirical studies have estimated a range of values of DTD but the general order of the size of the disutility is around 0.01 on average, which is equivalent to a loss of ~3.6 days of perfect health over one year. Although the cost-effectiveness of several primary preventative treatments has been shown to be sensitive to even small levels of DTD or treatment burden,[30-32] DTD values have only been elicited in a few small studies several of which have non-ideal sampling, and there is therefore considerable uncertainty as to the size and distribution of DTD values which limits their routine use in model-based CEA.[33,34]

Design: This study will use mixed methods comprising two choice-based stated preference surveys for statins and bisphosphonates whose design will be informed by qualitative semi-structured interviews. Four existing empirical studies have used a range of stated preference methods (time-trade off; standard gamble; willingness to pay) to elicit DTDs for a daily generic preventative pill,[33] a daily pill for cardiovascular prevention,[34,48] and aspirin or warfarin[49]. There is no agreement on the

appropriate method to use to elicit DTD[33,34,48,49] and so this study will identify and compare the values elicited from two stated preference methods: time-trade off and best-worst scaling.

1. *Elicitation study design phase*: The pilot scenarios to be used in the elicitation studies will be informed by the existing literature[33,34,48,49] and designed in collaboration with our public partners. The scenarios are (i) taking a statin daily for preventing CVD and (ii) taking a bisphosphonate weekly for preventing osteoporotic fragility fractures. The scenarios will then be used to pilot the elicitation studies with 15 current users of primary care, using cognitive interviews (think aloud) along with debriefing questions to refine the TTO and BWS survey format, structure and phrasing [13]. A quantitative pilot study with 50 respondents (recruited from ResearchNow and recruited from primary care) will be used to elicit some prior preferences to test the TTO and BWS design and also to inform the required sample size for the main survey (currently estimated ~1,000 respondents in total, 500 member of the general public and 500 with experience of taking relevant drugs recruited from general practice). Using data from all sources and again in collaboration with our public partners, the design of the web-based main-survey will then be finalised.

2. *Time-trade off elicitation study*: the time-trade off (TTO) method is a widely recognised standard method for eliciting utility values and has been used to generate a tariff of utility values for the EQ-5D-3 level recommended for use in technology appraisals for the National Institute for Health and Care Excellence (NICE).[50] A TTO study is designed around a hypothetical scenario in which a healthcare intervention or health state is described. In general, the TTO method involves asking respondents to consider the relative amounts of time (for example, number of life-years) they would be willing to sacrifice to avoid a poorer health state.[51] In this study, the TTO method will follow the approach taken by Hutchins et al[34,48] and ask respondents the maximum amount of time they are willing to give up at the end of their life to avoid having to take a medicine.

3. *Best-worst scaling elicitation study*: best-worst scaling experiments (BWS) are an extension of discrete choice experiments. BWS experiments ask respondents to select their most preferred and least preferred items (defined by attributes and levels) in a choice question.[52] An argued advantage of BWS over standard DCEs is that the choices made reveal more information about the relative strength of people's preferences for each attribute in the design using fewer questions which could, in turn, reduce the response error.[53] Importantly, using a BWS allows a rank ordering of the attributes in the experiment together with utility weights to be estimated. In this study, the BWS experiment will contain attributes and levels representing, for example, the duration and inconvenience of the treatment. Systematic reviews of the relevant published literature and input from the research team including patient involvement will be used to generate the list of potential attributes and levels. The semi-structured interviews in the design phase will be used to formulate the final choice and phrasing of the attributes and levels. Ngene software, applying Bayesian mathematical properties, will then be used to generate a pilot study design.

Sampling: Survey respondents will include the public and patients to elicit values from individuals without and with prior experience of taking a medicine long-term for a specific condition. Two samples of ~250 (for each case study: statin and bisphosphonate) of the general public will be recruited using an online panel company such as ResearchNow. ResearchNow provide access to respondent panels with pre-defined characteristics and can provide a demographically-balanced sample based on the characteristics of a particular area, in this case England. We have successfully used this recruitment strategy previously. Two samples of ~250 patients with experience of current or past use of taking a relevant long-term medicine (for each case study: statin and bisphosphonate) will be recruited from general practice. The total sample size for the main survey has been set around 1,000 respondents for now to estimate the necessary resources required to generate the sampling frame. This sample frame represents a reasonable estimate based on 250 responses per potential (n=4) sub-group and is in keeping with published guidance.[54,55]

Data collection: The TTO and BWS experiment will be embedded in a web-based survey hosted using Sawtooth software comprising four sections. The first section of the survey shows respondents training materials that, because the survey is on-line, can make use of short video clips or animation to explain the topic and purpose of the survey. Section two will contain the TTO. Section three will include the BWS. Section four will include questions collecting demographic data (for example: age; gender; currently taking a long-term prescribed medicine; presence of co-morbidities) that will be used to inform sub-groups to understand preference heterogeneity in the study sample. Different versions of the survey will be created in which the order of sections two and three are switched on a random basis to allow for analysis of ordering effects.

Data analysis: The cognitive interview data will be analysed using thematic analysis to inform key aspects of the survey design that need modifying. The aim of the quantitative data analysis is to estimate utility values to generate population and patient estimates for DTD to be used in study 3. The primary analysis will report the DTD for each case study estimated from TTO for the patients and the public samples. A secondary analysis will report the DTD estimated from BWS and compare the mean, range and distributions of the values elicited from patients and the public. Descriptive statistics (for TTO) and appropriate regression methods (based on conditional logit models for BWS) will be used to analyse the stated preference data.

Methods for objective 3: model-based cost-effectiveness analysis accounting for competing risks and direct treatment disutility

Objective 3: To examine the effect of accounting for competing risks and Direct Treatment Disutility on clinical effectiveness and relative cost-effectiveness in the context of the use of statins and bisphosphonates for the primary prevention of CVD and osteoporotic fracture respectively.

Rationale: Current NICE clinical guidelines make use of cost-effectiveness (CEA) evidence to inform recommendations about diagnosis and treatment. The ultimate goal for model-based CEA, is to compare the costs and benefits in order to promote (and recommend) interventions that provide the best value for money for the health service.[56] Preventative treatments have particular characteristics when compared with treatments for acute conditions and have both immediate and persistent harms but deferred benefits, with the risk of harm (as DTD) potentially greater in people with multimorbidity and high treatment burden. To date, there are few published examples of model-based CEA for preventative treatments that take account of, and quantify the impact of competing risks and DTD. Adapting current methods used to populate model-based cost-effectiveness analysis of preventative medicines is necessary to better reflect the potential for sub-groups of the patient populations to accrue different degrees of net benefit when the impact of DTD and/or competing risks are taken into account. Modifying existing methods in this way will allow cost-effective subgroups of patient populations to be defined that reflect different degrees of DTD (defined by specified levels of assumed DTD) and different levels of competing risk (defined by specified levels of competing risk at different levels of baseline risk) by identifying and quantifying the lifetime incremental costs and benefits (measured using QALYs) of preventative medicines compared with the alternative of no intervention.

Design: This study will use decision-analytic model-based cost-effectiveness analyses.

Sampling: The relevant patient population will be defined by the intervention for the two selected case studies. Each decision- analytic model-based CEA will use simulations for a hypothetical patient cohort. The size of the cohort defined will depend on the selected case study, but must be sufficiently large to allow for risk-based events to occur in the selected study time horizon indicating a suggested simulated cohort size of 200,000.[40]

Setting/context: This CEA will use two case studies to identify the incremental costs and benefits of preventative treatments taking into account the impact of competing risks and DTD for relevant

patient population taking (i) a statin for preventing CVD and (ii) a bisphosphonate for preventing osteoporotic fragility fractures.

Data collection: Two decision-analytic models will be structured and populated assuming the NHS study perspective and life-time horizon. The decision-analytic models used will be existing ones developed to inform NICE guidance produced for (i) a clinical guideline for lipid modifications[7] and a technology appraisal of bisphosphonates for preventing osteoporotic fragility fractures[40]. We have previously been granted access and have a fully executable version of the lipid guideline model, which is a Markov model. This model has been used by our research team in a previous project funded by NIHR HS-DR (11/2003/27 Better Guidelines project). One of the co-applicants on this current proposal (Davis) was the lead economist who developed the CEA model for the ongoing bisphosphonate technology appraisal. This technology appraisal is due for completion in November 2015 at which point the model will be released from embargo and made available to academic users under the standard NICE licence. This will provide the research team with access to a fully executable version of this discrete event simulation model. Both models will be populated with the original data (resource use, cost, clinical effectiveness, mortality, utility) used in the development of the respective clinical guideline and technology appraisal. Specific parameters (condition and all-cause mortality) will then be modified to account for the newly specified competing risks (from study 1). In addition, utility values will be modified to take account of the DTD values elicited in study 2. There is no standardised approach to use to adjust single event utility values for composite events[57] and so the base-case analysis will assume an 'additive' impact but scenario analyses will explore the impact of different assumptions.

Data analysis: The aim of the CEA is to use each case study to quantify the impact on incremental costs and benefits and of using the risk prediction model developed in study 1 to take account of competing risks and/or including DTD (generated in study 2) on absolute and relative QALYs, relative cost-effectiveness of statins and the relevant alternatives in this case of 'do nothing'. This will be done by creating patient vignettes to reflect clinically relevant examples, for each case study (CVD and osteoporotic fracture) of patients experiencing; (i) a DTD at different values; (ii) different levels of competing mortality risk to reflect competing risks of different amounts (iii) both DTD and different levels of competing mortality risk. This vignette based approach was previously used successfully in the Better Guidelines project.[28] Importantly, where appropriate, a range of scenario analyses will be used to quantify the impact of uncertainty on methodological assumptions, for example: use of different methods to generate the risk prediction model; use of different methods to elicit DTD or combine DTD with existing health states; impact of discount rate. In addition, if appropriate and feasible, given the model types, probabilistic sensitivity analysis will be used to capture the joint impact of parameter input values.

Dissemination and projected outputs

Expected output of research/impact

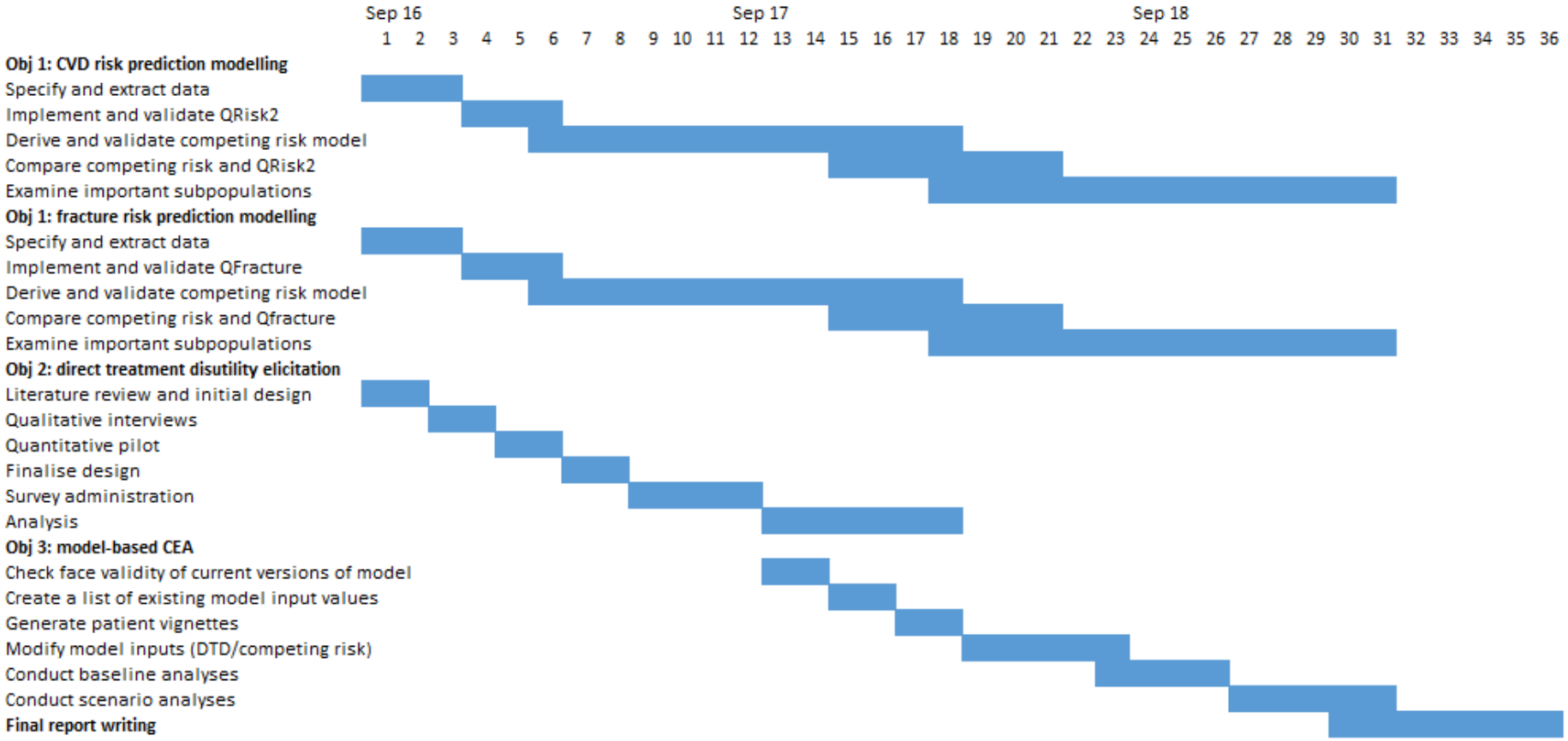
The final impact of the research will of course be driven by the findings of the study, and the extent to which new models provide better risk prediction or changed evidence of likely effectiveness and cost effectiveness of interventions. Our expectation is that the new risk prediction model will provide more accurate estimates of cardiovascular and fracture risk in the important subset of people with multimorbidity and high competing risks of death.[11] This has direct clinical implications in terms of avoiding low benefit or futile treatment. We also expect that the model-based cost-effectiveness analysis (CEA) will provide evidence on the relative cost-effectiveness in different groups of the population with different competing risks because of multimorbidity, exploring the important area of subgroup heterogeneity in cost-effectiveness analysis with implications for the creation of more appropriately nuanced guideline recommendations relating to preventative treatment.[27,29] Current CEA also do not consider the burden associated with taking a medicine long-term and there are few published estimates of the disutility associated with this treatment burden. This study will provide a

value of the impact of a direct treatment disutility (DTD) for a medicine taken long-term and also understand the impact of DTD on the relative cost-effectiveness of statins to prevent CVD and bisphosphonates to prevent fracture, thus taking account of patient and the public preferences for (not) taking a long-term treatment. Clinicians and patients already sometimes feel uncomfortable in using lifelong primary prevention in people with multimorbidity and reduced life expectancy, because of perceived uncertainty about whether average net benefit and therefore cost-effectiveness applies to them. This study will quantify net benefit and cost-effectiveness better and provide two improved risk prediction tools for use in clinical practice, and therefore contribute to improved guideline development and more patient-centred clinical decision making for a large number of people. More generally, the methods developed in relation to risk prediction accounting for competing risk of death, and model-based CEA accounting for DTD and in sub-populations with different competing risks will be more generally applicable in health technology assessment of, and guideline development for other long-term preventive interventions where individual benefit is relatively small and accrues over long periods.

Dissemination

Our plans for dissemination reflect that the findings will be of interest to multiple audiences, and will encompass both reporting of the findings of the individual case-studies, and methodological publication drawing on both case-studies to identify generalisable findings applicable to other preventative treatments or interventions. There will be academic, service-related and public outputs. Academic outputs will include formal reporting of the methods and findings in conference presentations, peer-reviewed publications in general medical and specialist medical and health economic journals, and the published peer-reviewed NIHR final report. We plan for conference presentations at the UK Health Economics' Study Group, the UK Society for Academic Primary Care conference, the Guidelines International Network conference, and the Medical Decision Making conference. While our academic outputs will report implications for clinicians and guideline developers, we will additionally produce appropriate summaries for clinical and guideline development audiences who are key users of the models we will examine, and offer workshops and seminars, for example to NICE and SIGN with whom we have established relationships (KP and BG have served on various NICE committees; BG chairs the NICE Multimorbidity guideline development group; our NIHR HS&DR Better Guidelines project was in collaboration with NICE and SIGN, and we presented the findings of it to the NICE Board and to the SIGN Strategy Board, and are working with technical teams on implementation). Specifically, we will organise two half-day workshop-style conferences separately targeting an academic and technical audience (including guideline development technical teams and academics with an interest in prognostic modelling and reporting standards for such studies [eg http://progress-partnership.org/]) and a clinical/primary care IT/public audience including major GP clinical system suppliers and national charities or advocacy groups (or a one day conference with the option to attend one or both half days). The academic/technical workshop will examine the implications of competing risks and direct treatment disutility for risk prediction and model-based CEA, whereas the clinical/primary care IT/public workshop will focus on the implications for clinical and shared decision making, and how clinical IT systems and web-based resources can robustly support that. As well as making available web-based risk prediction calculators for clinical and public use (with suitable guidance for interpretation), we will publish our full algorithms to support external validation by others, and in a form which will facilitate use of the algorithms in clinical care, for example by embedding them in clinical IT systems and will initiate discussions with system suppliers at the additional planned workshop/conference. We will publish our definitions and codesets for defining morbidities at ClinicalCodes.org (https://clinicalcodes.rss.mhs.man.ac.uk/) to facilitate reuse by other researchers. Working with our public partners, we will also produce suitable lay summaries and disseminate key findings via general press and media.

Figure 1: Gantt chart



Plan of investigation and timetable

The Gantt chart in figure 1 above shows the timing of the different elements of the project. We will obtain approval for the objective 1 statistical modelling from the CPRD Independent Scientific Advisory Committee before the project officially starts. Ethics approval will be obtained by month 2 before public and patient recruitment to the objective 2 elicitation study.

Project management

The study will be managed across two main sites (Dundee and Manchester) with input from one co-applicant (SD) based in Sheffield. BG is the overall chief investigator and will be responsible for the timely and rigorous conduct of the whole project. He will lead and be responsible for delivering the Dundee-based objective 1 work (competing risks modelling) with the support of DM and PD for data specification, extraction, management and analysis. KP will lead and be responsible for delivering the Manchester-based objectives 2 and 3 work (direct treatment disutility elicitation and economic modelling) which AT will co-ordinate and for which SD will have an advisory role, particularly for adaptation of the NICE bisphosphonates model which she developed. We have successfully run a previous NIHR HS&DR project using a similar split of work between Dundee and Manchester (NIHR 11/2003/27 Better Guidelines grant), so have experience of working together. Each site will have weekly to fortnightly internal meetings, and the overall Study Management Group consisting of the co-applicants will have regular meetings (2-6 weekly depending on the stage of the project) which will predominately be held by videoconference, but with ~12 weekly face-to-face longer meetings.

We will convene a Project Advisory Group (PAG) consisting of the co-applicants plus our two public partners plus relevant experts which will meet 3-4 times over the course of the project with additional e-mail or tele/videoconference contact as required. The role of the PAG is to provide advice and guidance to the Study Management Group on the design of the study and the interpretation of the results, including on how best to disseminate them to different audiences. We will invite appropriate stakeholders to join the PAG including an economist and epidemiologist with relevant expertise, and professionals with guideline development expertise (Dr Gerry Richardson, Senior Research Fellow in the Team for Economic Evaluation and Health Technology Assessment at the University of York, and Dr Nichole Taske, Associate Director for Methodology at NICE have already agreed to join).

We will separately convene an independent Study Steering Committee (SSC) consistent with NIHR HS&DR guidance. The SSC will have responsibility for overall project governance with majority independent membership including independent chair, economist, statistician/epidemiologist and two public members (additional to those on the PAG). This will meet 3-4 times over the course of the project.

Approval by ethics committees

There are no major ethical issues raised. The objective one modelling using CPRD data will be subject to the approval of the CPRD Independent Scientific Advisory Committee and will follow best practice for the management and storage of anonymised data. The objective two elicitation work will require NHS Research Ethics Service or University Ethics Committee review and approval, but given the low risk nature of the research, we do not envisage that this will pose any problems. The objective three modelling uses existing NICE economic models which are available for this purpose subject to NICE's standard licence conditions. We already have the NICE lipid modification model, and one of the co-applicants (SD) developed the NICE bisphosphonate model. We will apply for and obtain ISAC approval before the study starts, and will apply for NHS RES approval before the study starts completing full review before month two.

Patient and Public Involvement

In preparing this application, we drew on qualitative analysis of patient data from 8 focus groups with 48 participants and 9 individual interviews about prescribing and prescribing decision-making which we did as part of our DQIP prescribing safety improvement programme which finished at the end of 2014. We additionally carried out a group discussion with 8 members of an NHS Patient and Public Participation group, and our two public partners contributed to the development of the proposal through discussion and through their membership of our previous NIHR HS&DR funded 'Better Guidelines' project reference group. Unsurprisingly, across these groups, there was strong support for the idea that treatment decisions for individual conditions should take account of other conditions, other treatments and the context of the individual as a whole person, and for work examining whether such accounting could improve the quality of evidence which underpinned treatment decisions. This was significantly driven by a general perception that the number of drugs people were taking regularly was increasing, and some unease about whether the benefits of this always outweighed the harms.

The two public members from our NIHR funded Better Guidelines project (Graham Bell and Alison Allen) have agreed to join the advisory group, and provided input to the plain English summary. The proposed modelling is highly technical and in discussion with our public partners, we have agreed that for this element, their participation will be in oversight of progress and contribution to interpretation and dissemination through advisory group involvement rather than direct involvement in model design and implementation. The public partners will have a more direct role in the design of the Direct Treatment Disutility study, which involves eliciting utilities from members of the public and patients, and where careful design of the participant information material and study processes is critical. Given the technical complexity of the methods, we will offer training in the broad methods used, and discussion of advisory group papers before meetings if required. As well as paying direct expenses, we will offer an honorarium for attendance at meetings and preparation time at the INVOLVE rate for committee meetings, and similarly for support in developing study materials.

Expertise and justification of support required

BG is an academic GP and health services researcher with an interest in prescribing, multimorbidity and guidelines, and in-depth experience of statistical modelling using GP clinical data and other linked datasets similar to the data that will be analysed in Dundee. He co-led (with KP and in collaboration with NICE and SIGN) the recently completed NIHR HS&DR funded 'Better Guidelines' project, which examined how existing single disease guidelines could better account for multimorbidity. He is a member of a number of NHS committees and chairs the guideline development group for the NICE Multimorbidity guideline which will publish in autumn 2016. Costs are requested for 15% of his time. DM is a GP and pharmacoepidemiologist, with experience of obtaining ISAC approval, specifying and extracting data from CPRD, and analysis of large quantitative datasets including statistical modelling. His current work includes leading a project using CPRD data to compare different methods of predicting mortality in people with COPD. PD is Professor of Epidemiology and Biostatistics in Dundee with expertise in analysis of routine data and experience of developing risk prediction models including for cardiovascular disease in people with type 2 diabetes and for emergency hospital admission in older people. Costs are requested for 5% of DM and PDs' time.

KP is Professor of Health Economics in Manchester, with an interest in health technology assessment and led the health economics work-stream of the Better Guidelines project, which included adapting existing NICE cost-effectiveness models to examine how lifetime absolute QALY gain varied after accounting for plausible levels of direct treatment disutility and competing risk. AT is a health economist who was the employed researcher on the Better Guidelines project health economic work-stream and therefore has experience in the model adaptations required. SD is Senior Lecturer in Health Economics at the University of Sheffield with expertise in modelling the cost-effectiveness of treatments for osteoporosis, and will specifically provide advice on adaptation of the NICE

bisphosphonates model which she led the development of. Costs are requested for 15% of AT's and 5% of KP and SDs' time.

Our two public partners (Alison Allen and Graham Bell) have experience of working the Scottish Intercollegiate Guidelines Network, and with us on our previous guidelines project, and will be public partners on the project advisory group (PAG). Costs are requested at the INVOLVE day rate for their participation in the PAG (4 days) and their support for the objective 2 elicitation study materials design (2 days) plus travel expenses. Additionally, Dr Gerry Richardson, Senior Research Fellow in the Team for Economic Evaluation and Health Technology Assessment at the University of York and Dr Nichole Taske, Associate Director (Methodology) at NICE have already agreed to be members of the PAG.

The models being examined are central to both clinical and guideline development/health technology assessment decision making in relation to treatments for primary prevention. Many of the most commonly prescribed drugs are for primary prevention including most prescriptions for statins and bisphosphonates, so the consequences of these decisions affect very large numbers of people, have very large aggregate cost, and have contributed to increases in polypharmacy[2] and treatment burden[4]. Appropriately targeted primary prevention is cost-effective and fits with the preferences of most people at risk, but the appropriateness of aggressive primary prevention in people with multimorbidity and reduced life expectancy is less certain, not least because existing risk prediction tools either ignore competing risk (QRisk2, QFracture) or incompletely account for it (eg the QRisk lifetime model which does account for competing risk but estimates risk of non-CVD death only in terms of patient characteristics predicting CVD events, which effectively ignores the specific effects of an individual having non-CVD life-limiting conditions). The research therefore has the potential to make guideline recommendations more properly nuanced to account for individual circumstances, to make clinical decision making more patient centred, and to improve the value for money of primary prevention by targeting treatment on those most likely to benefit.

Funding

The study is funded by the NIHR Health Services and Delivery Research Programme (NIHR HS&DR 15/12/22).

References

1. Barnett K, Mercer SW, Norbury M et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. The Lancet 2012;380:37-43.
2. Guthrie B, Makubate B, Hernandez-Santiago V, Dreischulte T. The rising tide of polypharmacy and drug-drug interactions: population database analysis 1995-2010. BMC Medicine 2015;13:74.
3. Dumbreck S, Flynn A, Nairn M, et al. Drug-disease and drug-drug interactions: systematic examination of recommendations in 12 UK national clinical guidelines. BMJ 2015;350:h949.
4. May C, Montori VM, Mair FS. We need minimally disruptive medicine. BMJ 2010;339.
5. Hughes L, McMurdo MET, Guthrie B. Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity. Age and Ageing 2013;42:62-9.
6. Guthrie B, Payne K, Alderson P, McMurdo MET, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. BMJ 2012;345:e6341.
7. National Institute for Health and Care Excellence. Clinical Guideline 181: Lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. London: National Institute for Health and Care Excellence; 2014.
8. National Institute for Health and Care Excellence. Short clinical guideline CG146 - Osteoporosis: fragility fracture risk. London: National Institute for Health and Care Excellence; 2012.
9. National Osteoporosis Guideline Group. Osteoporosis: clinical guideline for prevention and treatment. Sheffield: National Osteoporosis Guideline Group; 2014.

10. Bisphosphonates for preventing osteoporotic fragility fractures (including a partial update of NICE technology appraisal guidance 160 and 161, expected publication November 2015). https://www.nice.org.uk/guidance/indevelopment/gid-tag462. 2015.
11. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? Statistics in Medicine 2012;31:1089-97.
12. Wolkewitz M, Cooper BS, Bonten MJM, Barnett AG, Schumacher M. Interpreting and comparing risks in the presence of competing events. BMJ 2014;349:g5060.
13. Wolbers M, Koller MT, Stel VS, et al. Competing risks analyses: objectives and approaches. Eur Heart J 2014;35:2936-41.
14. Pintilie M. Analysing and interpreting competing risk data. Statistics in Medicine 2007;26:1360-7.
15. Maki E. Power and sample size considerations in clinical trials with competing risk endpoints. Pharm Stat 2006;5:159 - 71.
16. Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic Models With Competing Risks: Methods and Application to Coronary Risk Prediction. Epidemiology 2009;20:555-61.
17. Koller MT, Leening MJG, Wolbers M, et al. Development and Validation of a Coronary Risk Prediction Model for Older U.S. and European Persons in the Cardiovascular Health Study and the Rotterdam Study. Annals of Internal Medicine 2012;157:389-97.
18. van Staa T-P, Gulliford M, Ng ESW, Goldacre B, Smeeth L. Prediction of Cardiovascular Risk Using Framingham, ASSIGN and QRISK2: How Well Do They Predict Individual Rather than Population Risk? PLoS ONE 2014;9:e106455.
19. Hippisley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. 2010;341:c6624.
20. Ferket BS, van Kempen BJH, Heeringa J, et al. Personalized Prediction of Lifetime Benefits with Statin Therapy for Asymptomatic Individuals: A Modeling Study. PLoS Med 2012;9:e1001361.
21. Bourgeois FT, Shannon MW, Valim C, Mandl KD. Adverse drug events in the outpatient setting: an 11-year national analysis. Pharmacoepidemiology and Drug Safety 2010;19:901-10.
22. Kutner JS, Blatchford PJ, Taylor DH, Jr, et al. Safety and benefit of discontinuing statin therapy in the setting of advanced, life-limiting illness: A randomized clinical trial. JAMA Internal Medicine 2015;175:691-700.
23. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. 2008;336:1475-82.
24. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study2012.
25. House of Commons Health Committee. Managing the care of people with long–term conditions http://www.parliament.uk/business/committees/committees-a-z/commons-select/health-committee/news/long-term-conditions-substantive/. London: House of Commons; 2014.
26. Rawlins M. Six hopes and dreams. National Institute for Care Excellence. Birmingham2012.
27. Sculpher M. Subgroups and Heterogeneity in Cost-Effectiveness Analysis. PharmacoEconomics 2008;26:799-806.
28. Guthrie B, Thompson A, Dumbreck S, et al. Better guidelines for better care: accounting for multimorbidity in clinical guidelines (NIHR HS&DR 11/2003/27). Draft final report submitted June 20152015.
29. Espinoza MA, Manca A, Claxton K, Sculpher MJ. The Value of Heterogeneity for Cost-Effectiveness Subgroup Analysis: Conceptual Framework and Application. Medical Decision Making 2014;34:951-64.
30. Lazar LD, Pletcher MJ, Coxson PG, Bibbins-Domingo K, Goldman L. Cost-Effectiveness of Statin Therapy for Primary Prevention in a Low-Cost Statin Era. Circulation 2011;124:146-53.
31. Vijan S, Sussman JB, Yudkin JS, Hayward RA. Effect of patients' risks and preferences on health gains with plasma glucose level lowering in type 2 diabetes mellitus. JAMA Internal Medicine 2014.
32. Timbie JW, Hayward RA, Vijan S. Variation in the Net Benefit of Aggressive Cardiovascular Risk Factor Control Across the US Population of Patients With Diabetes Mellitus. Arch Intern Med 2010;170:1037-44.
33. Fontana M, Asaria P, Moraldo M, et al. Patient-Accessible Tool for Shared Decision Making in Cardiovascular Primary Prevention: Balancing Longevity Benefits Against Medication Disutility. Circulation 2014;129:2539-46.

34. Hutchins R, Pignone MP, Sheridan SL, Viera AJ. Quantifying the utility of taking pills for preventing adverse health outcomes: a cross-sectional survey. BMJ Open 2015;5.
35. Herrett E, Shah A, Boggon R, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ 2013;346.
36. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. British Journal of Clinical Pharmacology 2010;69:4-14.
37. Khan N, Harrison S, Rose P. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract 2010;60:e128-36.
38. Hippisley-Cox J, Coupland C, Brindle P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. BMJ Open 2014;4.
39. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. Journal of Clinical Epidemiology 2003;56:441-7.
40. Davis S, Martyn-St James M, Sanderson J, et al. Bisphosphonates for preventing osteoporotic fragility fractures (including a partial update of NICE technology appraisal guidance 160 and 161). Technology Assessment Report: Final report to the National Institute for Health and Care Excellence2015.
41. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age and Ageing 2016;Advance access doi: 10.1093/ageing/afw039.
42. Greving J, Visseren F, de Wit G, Algra A. Statin treatment for primary prevention of vascular disease: whom to treat? Cost-effectiveness analysis2011.
43. Pignone M, Earnshaw S, Tice JA, Pletcher MJ. Aspirin, Statins, or Both Drugs for the Primary Prevention of Coronary Heart Disease Events in Men: A Cost–Utility Analysis. Annals of Internal Medicine 2006;144:326-36.
44. Pletcher MJ, Pignone M, Earnshaw S, et al. Using the Coronary Artery Calcium Score to Guide Statin Therapy: A Cost-Effectiveness Analysis. Circulation: Cardiovascular Quality and Outcomes 2014;7:276-84.
45. Pletcher MJ, Lazar L, Bibbins-Domingo K, et al. Comparing Impact and Cost-Effectiveness of Primary Prevention Strategies for Lipid-Lowering. Annals of Internal Medicine 2009;150:243-54.
46. Roberts E, Horne A, Martin S, Blaha M, Blankstein R, Budoff M. Cost-Effectiveness of Coronary Artery Calcium Testing for Coronary Heart and Cardiovascular Disease Risk Prediction to Guide Statin Allocation: The Multi- Ethnic Study of Atherosclerosis (MESA). PLoS One 2015;10.
47. Thompson A, Guthrie B, Payne K. Do pills have no ills: capturing the impact of direct treatment disutility? . Pharmacoeconomics (under review) 2015.
48. Hutchins R, Viera AJ, Sheridan SL, Pignone MP. Quantifying the Utility of Taking Pills for Cardiovascular Prevention. Circulation: Cardiovascular Quality and Outcomes 2015;8:155-63.
49. Gage BF, Cardinalli AB, Owens DK. THe effect of stroke and stroke prophylaxis with aspirin or warfarin on quality of life. Archives of Internal Medicine 1996;156:1829-36.
50. National Institute for Health and Care Excellence. Methods guide for technology appraisal. London: National Institute for Health and Care Excellence; 2013.
51. Torrance G. Measurement of health state utilities for economic appraisal. J Health Econ 1986 1986;5:1-30.
52. Flynn T, Louviere J, Peters T, Coast J. Best–worst scaling: what it can do for health care research and how to do it. J Health Econ 2007;26:171–89.
53. Xie F, Pullenayegum E, Gaebel K, Oppe M, Krabbe P. Eliciting preferences to the EQ-5D-5L health states: discrete choice experiment or multiprofile case of best-worst scaling? Eur J Heal Econ 2014;15:281–8.
54. de Bekker-Grob E, Donkers B, Jonker M, Stolk E. Sample size requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. Patient 2015;DOI 10.1007/s40271-015-0118-z.
55. Johnson Fea. Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value in Health 2013;16:3-13.
56. Wonderling D, Sawyer L, Fenu E, Lovibond K, Laramée P. National Clinical Guideline Centre Cost-Effectiveness Assessment for the National Institute for Health and Clinical Excellence. Annals of Internal Medicine 2011;154:758-65.

57. Ara R, Wailoo A. NICE Decision Support Unit technical support document 12: the use of health state utilitiy values in decision models http://www.nicedsu.org.uk/TSD12%20Utilities%20in%20modelling%20FINAL.pdf. London: National Institute for Health and Care Excellence; 2011.