A Factored Co-morbidity Approach for Modelling CKD Progression

Norman Poh[1], Simon Bull[1], Santosh Tirunagari[1], Christopher Farmer[2], Nicholas Cole[3] and Simon de Lusignan[3]

1: Department of Computer Science, University of Surrey

2: East Kent Hospitals University NHS Foundation Trust

3: Department of Clinical and Experimental Medicine, University of Surrey

Contact: N.Poh@surrey.ac.uk

Introduction

One of the key challenges in understanding CKD progression is its multifaceted aetiology. This is evident as it is commonly observed that hypertension, anaemia, heart disease and diabetes are common co-morbidities of CKD. In other words, the existence of co-morbidities can potentially alter the risk of CKD progression, e.g. from stage 3 to 5. Similarly, treatment for one disease may be detrimental to kidneys. Unfortunately, "flat" risk models such as logistic regression, e.g. as implemented by the QKidney score and many similar risk models, are not designed to extract the rich structure induced by a multitude of co-morbidities, the state of which are often captured in routinely collected patient data.

We revisit the modelling of CKD progression by proposing a Bayesian network that has a certain regular structure, which has attractive properties such as tractability and efficient computation. The risk of progression to more severe CKD stages can be inferred not only from eGFR but signs and symptoms (s) related to kidney impairment, while causality of CKD is captured by risk factors (r) and treatment strategies (t).

CKD questions

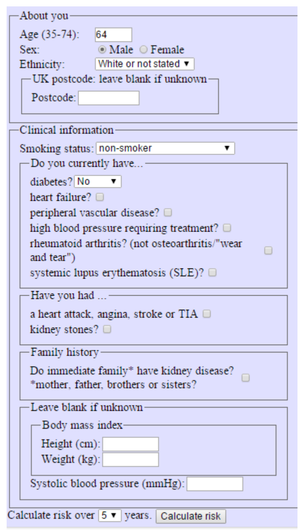
1. **Case finding**: typically used for calculating CKD prevalence. The detection accuracy differs depend on context (primary care, community care, someone who turns up in A&E?)
2. **Estimating CKD risk**: For screening a population with yet undiagnosed CKD, typically applied to population at risk of CKD, e.g., diabetes.
3. **Determining CKD stages**: applied to a population who have confirmed CKD. Current solution: KDIGO guidelines
4. **Estimating CKD progression risk**:
   - In routine follow-up of mild CKD patients (stages 1 and 2) in primary care. Specifically, progression into stage 3 triggers clinical interventions
   - Follow-up of CKD patients (stages 1—3) who progress into stages 4&5

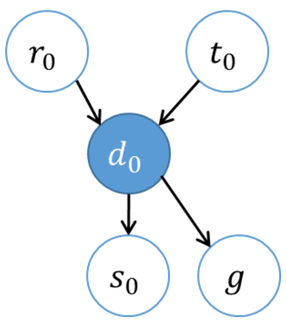
Bayesian networks

List of variables

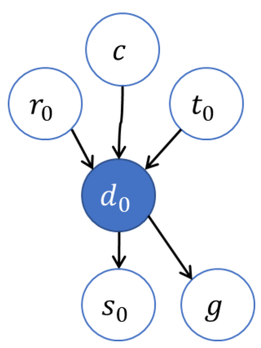
Variable	Meaning
c	Covariates, e.g., age, sex, index of multiple deprivation
r_0	risk factors, heart attack, angina, stroke or TIA; kidney stones; Family history of CKD; history of kidney impairments
t_0	Treatments for CKD
s_0	Signs and symptoms for CKD, kidney impairments
g	eGFR (estimated Glomerular Filtration Rate)
g'	eGFR slope (annual rate change)
m_0	Laboratory measurements known to covary with CKD, phosphate
d_0	disease diagnosis related to CKD, e.g., presence/absence of CKD and/or its stages
d_i	Each i indicates one CKD comorbidity, which can be a cause or through association. E.g., diabetes, heart failure, rheumatoid arthritis, hypertension, anaemia, CVD, PVD, etc.

Qkidney score (http://www.qkidney.org)

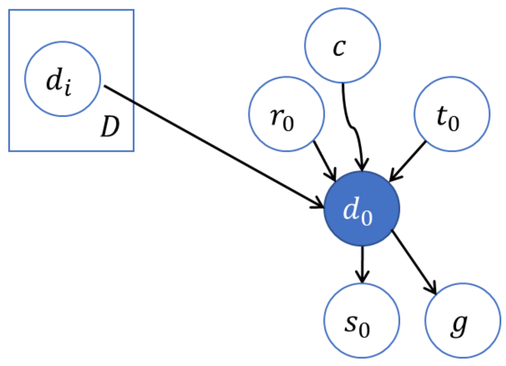




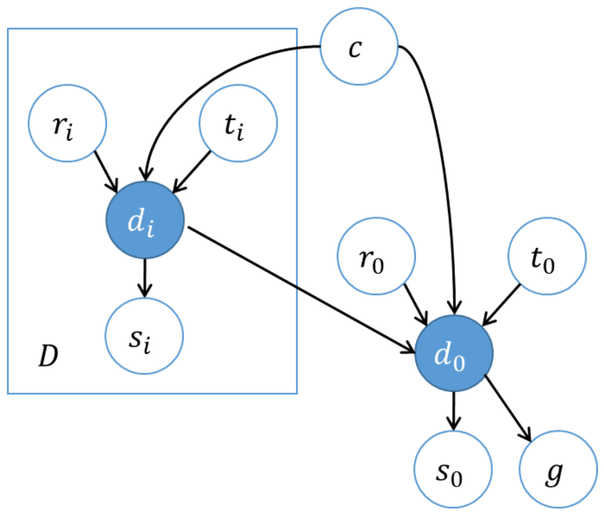
(a) Basic model



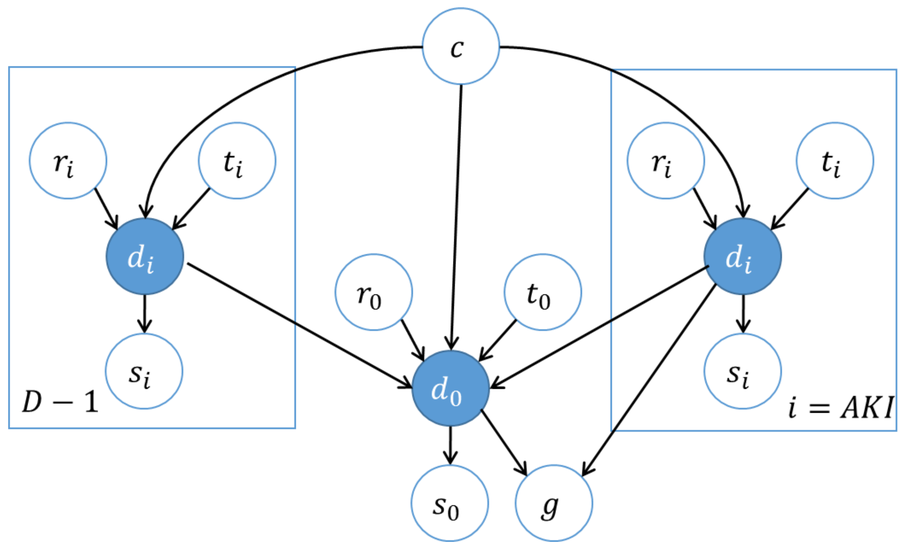
(b) Basic model with covariates



(c) Model with comorbid diagnosis



(d) General model with unobserved comorbid diagnosis



(e) Modified model for predicting CKD patients considering the risks of commonly associated co-morbidities, including acute kidney injury

Clinical relevance

We consider three contexts

1. Extracted data: carefully curated and prepared data, which are typical in clinical trials and register data
2. Clinical/primary care data: routinely collected coded data from clinics/GPs which are encounter-based
3. Hospital/secondary care data: episode-based data, coded in ICD

Results: Formulated questions

1. Is this a CKD patient?

Context	Input variables	Output variables	Algorithm
Extracted data	$\{r_0, t_0, s_0, m_0, d_0, g\}$	Binary output: CKD=yes, CKD=no	• Deterministic: Rules based
Primary care	Read codes of $\{r_0, t_0, s_0, d_0\}$, and $\{m_0, g\}$	Binary output: CKD=yes, CKD=no	• Deterministic: Rules based • Probabilistic: Logistic regression
Secondary care	ICD10 codes of $\{r_0, t_0, s_0, d_0\}$, and $\{m_0, g\}$	Binary output: CKD=yes, CKD=no	• Deterministic: Rules based • Probabilistic: Logistic regression

2. Given a non-CKD patient, what is her risk of contracting CKD?

Context	Input variables	Output variables	Algorithm
Extracted data	$\{r_0, t_0, s_0, m_0, d_i \mid \forall_i\}$ plus covariates, c	Binary output: CKD=yes, CKD=no	• Probabilistic: Logistic regression
Primary care	Read codes of $\{r_0, t_0, s_0, d_i \mid \forall_i\}$, and $\{m_0, g\}$ plus covariates, c	Binary output: CKD=yes, CKD=no	• Probabilistic: Logistic regression
Secondary care	ICD10 codes of $\{r_0, t_0, s_0, d_i \mid \forall_i\}$, and $\{m_0, g\}$ plus covariates, c	Binary output: CKD=yes, CKD=no	• Probabilistic: Logistic regression

3. Given a CKD patient, what is her CKD stage?

Context	Input variables	Output variables	Algorithm
Extracted data	Kidney impairments r_0 and recent eGFR values: $\{g\}$	Ordinal output: CKD stages 1—5	• Deterministic: KDIGO
Primary care	Read codes of kidney impairments and $\{r_0\}$ and recent eGFR values: $\{g\}$	Ordinal output: CKD stages 1—5	• Deterministic: KDIGO • Probabilistic approach: broken sticks model
Secondary care	kidney impairments and $\{r_0\}$ and recent eGFR values: $\{g\}$	Ordinal output: CKD stages 1—5	• Deterministic: KDIGO • Probabilistic approach: broken sticks model

4. Given a CKD patient and her existing CKD stage, what is her future stage?

Context	Input variables	Output variables	Algorithm
Extracted data	Variables in Qkidney score	Binary output: progress to severe CKD stages or not	• Logistic regression
Primary care	Read codes of $\{r_0, t_0, s_0, d_0\}$, and $\{m_0, g\}$ plus covariates plus eGFR slope	Binary output: progress to severe CKD stages or not	• Logistic regression and more advanced models
Secondary care	$\{r_0, t_0, s_0, d_0\}$, and $\{m_0, g\}$ plus covariates plus eGFR slope	Binary output: progress to severe CKD stages or not	• Logistic regression and more advanced models

Discussion

- Current risk models are 'flat' and do not exploit the rich information captured in patient records. The proposed factored co-morbidity approach offers a natural next step whilst being tractable and efficient in computation.
- Not all clinical questions have been systematically addressed using computational models
- The merit of the proposed approach will be investigated as part of the MRC Modelling CKD project using data extracted from the Royal College of GPs database.

www.modellingCKD.org

Acknowledgement: This project entitled "Modelling the Progression of CKD" was awarded to NP by the Medical Research Council, under the New Investigator Grant Scheme MR/M023281/1.