# Initialising & Calibrating Prevalence in the CKD Model

#### Version 2: 29 May 2025

## Introduction

Explainer and detail of how prevalence figures were derived to initialise and calibrate the CKD model. Whilst diagnosed prevalence has been recorded and reported reasonably well for some time, due to QOF register and now the CVD Prevent audits, undiagnosed prevalence has been harder to define and quantify. Published papers of previous models (some of which have been used in cross-validation) are vague in how they have calculated counts or rates of undiagnosed, and so there are large discrepancies between models.

## Context

A brief summary of how other models/papers have tackled undiagnosed:

* ‘PHE CKD prevalence model’ (2014) uses HSE 2012 data - due to the nature of how the survey was conducted, this used a single low eGFR to define someone as having CKD.
* ‘Impact CKD’ (2024) and ‘UKRR a public health emergency’ (2023) referred to HSE 2016 figures – due to the nature of how the survey was conducted, this used a single low eGFR to define someone as having CKD.
* ‘Inside CKD’ (2024) used 5+ years of linked CPRD-HES data to derive rates per hundred thousand of stage and diagnosis status (but did not define what constitutes as ‘undiagnosed’). The paper also references the PHE (2014) model.

THE PHE and UKRR models do not explicitly break down prevalence figures into diagnosed and undiagnosed.

## CVD Prevent

CVD Prevent have been routinely extracting GP data relating to cardiovascular health on a quarterly basis since March 2022 (plus one data point in September 2021). Of particular interest to this project are the following indicators:

* CVDP001CKD: Prevalence of GP recorded CKD (G3a to G5)
* CVDP002CKD: Uncoded – two low eGFRs with no recorded CKD
* CVDP003CKD: High risk – one low eGFR with no recorded CKD (the metadata specifically mentions that patients counted here are not included in ‘Uncoded’, so there is no risk of double-counting occurring).

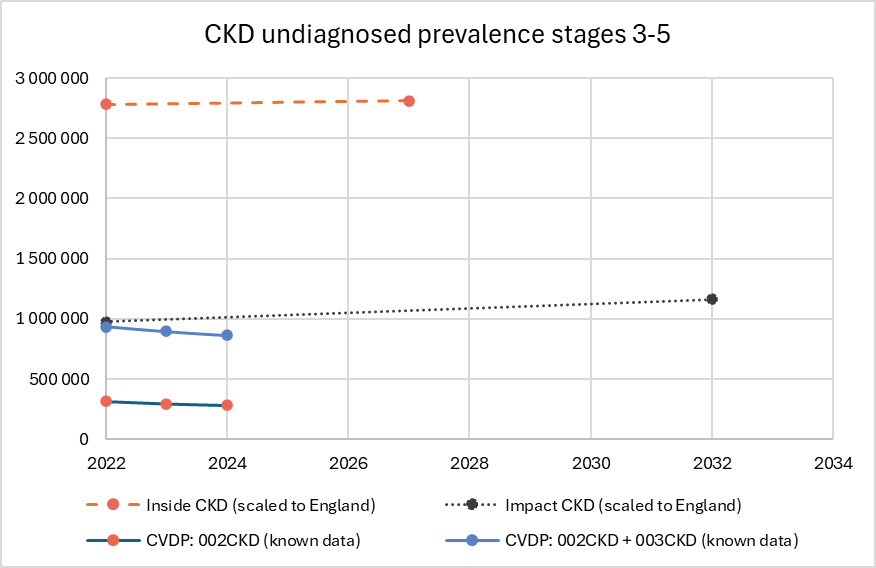
Indicator 1 is used as the count of diagnosed.

Indicator 3 seems to correlate with the HSE definition of undiagnosed, of using the results of just one (the latest) eGFR. Indicator 2 uses the more rigorous test results that define CKD, but note that they haven’t been recorded by GP as having CKD.

## Comparison

### Undiagnosed

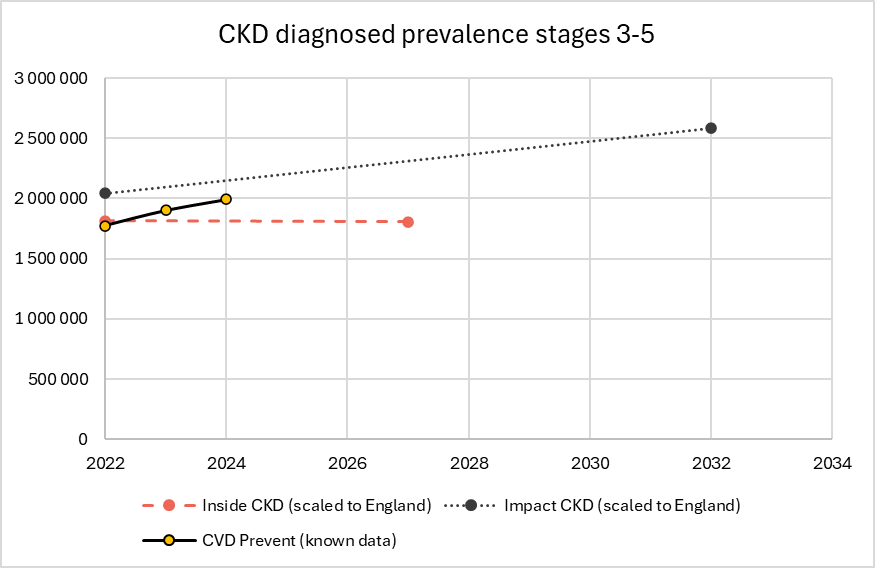
For the counts of undiagnosed, a decision was needed as to whether to use Indicator 2 (uncoded), or the sum of Indicators 2 & 3 (uncoded plus high risk). Both options were compared with other models to help inform the decision.



Using the sum of indicators 2 & 3 as undiagnosed aligns closely with the start value used in ‘Impact CKD’ model, and highlights the extreme difference between Impact CKD and Inside CKD.

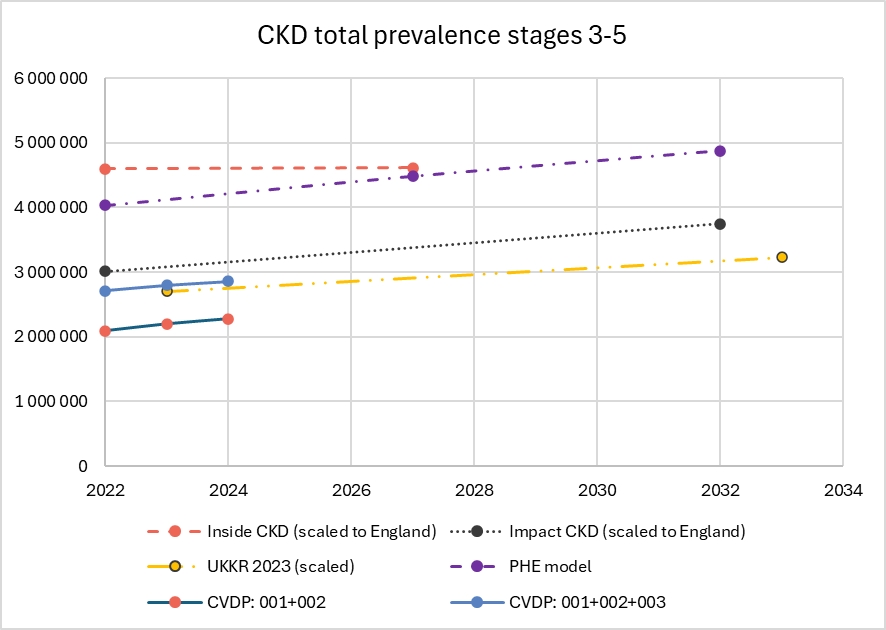
### Diagnosed

The diagnosed value from CVD Prevent aligns closely with that used in ‘Inside CKD’. ‘Impact CKD’ has a slightly higher count of diagnosed.



### Combined (total) prevalence

Using the sum of all 3 indicators for total prevalence aligns closely with that used in the UKRR model, and with the ‘Impact CKD’ model (once the higher value for diagnosed is taken into account).



On the basis of this comparison, it was decided to use the data from CVD Prevent for all prevalence counts, with undiagnosed being the sum of sum of indicators 2 & 3. This also provides more internal consistency, as there are 13 data points with which to calibrate the model against.

## Prevalence by CKD Stage

CVD Prevent does not report the CKD stage, and so methods from other models had to be considered to determine counts by stage and diagnosis state.

‘Inside CKD’ made use of linked CPRD-HES data to determine rates by stage and diagnosis state for a UK population. From this, it was possible to determine, for each diagnosis state, the percentage that were stage 3.

|  |  |
| --- | --- |
| **Diagnosis state** | **S3 as % of total state** |
| *overall* | *94.9%* |
| diagnosed | **90.1%** |
| undiagnosed | **98.0%** |

These figures were applied to the prevalence counts from CVD Prevent for each reporting period to derive counts by diagnosis and stage for use in calibrating the CKD SD model.

(CVD Prevent data was available to end of 2024, but a change in reporting or definitions caused a change in levels in 2024 Q3 onwards, which negatively affected the calibration).

## Calibration

The model was calibrated, using optimisation, to determine the optimal flow (per hundred thousand) between stages and states such that the model matched the known data points. In the screenshot below, the red line indicates the CVD Prevent values and the blue line is the model output.

The data used in the calibration were the quarterly reporting periods between March 2022 and June 2024.

A graph of a function

AI-generated content may be incorrect.

The optimisation process determined values for the flows between each stock (stage and state). These values (in people per hundred thousand of the upstream stock) were then fixed for the duration of the simulation.