

# Counting drugs to understand the disease: The case of measuring the diabetes epidemic

Henrik Støvring<sup>\*1</sup>, Morten Andersen<sup>1</sup>, Henning Beck-Nielsen<sup>2</sup>, Anders Green<sup>3</sup> and Werner Vach<sup>4</sup>

<sup>1</sup>Research Unit of General Practice, University of SouthernDenmark, J.B. Winsløvsvej 9A, 5000 Odense C, Denmark

<sup>2</sup>Diabetes Research Center, Odense University Hospital, Klørvænget 6<sup>4</sup>, 5000 Odense C, Denmark

<sup>3</sup>Department of Epidemiology, J.B. Winsløvsvej 9B, 5000Odense C, University of Odense, Denmark

<sup>4</sup>Department of Statistics, University of SouthernDenmark, Campusvej 55, 5230 Odense, Denmark

Email: Henrik Støvring\* - hstovring@health.sdu.dk; Morten Andersen - mandersen@health.sdu.dk; Henning Beck-Nielsen - henning.beck-nielsen@ouh.fyns-amt.dk; Anders Green - agreen@health.sdu.dk; Werner Vach - wvach@stat.sdu.dk;

\*Corresponding author

## Abstract

---

**Background:** Diabetes prevalence increases globally with severe consequences for afflicted individuals and societies. Data on diabetes incidence and diabetes related mortality on a population level are, however, scarce. As an alternative to dedicated studies it has been suggested to use pharmacoepidemiological databases that are readily available, at least in the Nordic countries.

**Methods:** For all 470,000 inhabitants in Funen County, Denmark, in the period 1992-2003, data on gender, date of birth, death and migration to and from the county, and any filled prescriptions of an anti-diabetic medication was obtained from the Odense Pharmaco-Epidemiological Database.

**Results:** Prevalence odds for use of an anti-diabetic medication rose annually 3.5% (95% confidence interval: 3.1%, 3.9%) for females, 4.5% (4.0%, 4.9%) for males. Corresponding incidence rates annually rose 4.8% (3.8%, 5.8%) for females, 4.5% (3.5%, 5.4%) for males. Mortality rates among treated annually declined 2.8% (1.4%, 4.1%) among females, 2.2% (0.9%, 3.5%) among males. The disequilibrium in absolute numbers between incidence and mortality among treated was the main driver for the increasing prevalence, while concurrent trends in incidence and diabetes related mortality only marginally affected prevalence trends. Trend estimates were insensitive to varying the length of the run-in period used for determining treatment status, except when using the naive and methodologically flawed run-in period of variable length.

**Conclusions:** While pharmacoepidemiological databases provide a useful tool for monitoring pharmacologically treated diabetes, a dedicated diabetes database covering all prevalents and incidents is needed for a more detailed analysis of underlying causes and trends.

---

## Background

The population epidemiology of diabetes mellitus is of public health interest from several perspectives. Diabetes has severe costs for afflicted patients, as it is a chronic lifelong disease characterized by excess mortality [1] and high comorbidity with cardiovascular disease, nephropathy and eye complications as prominent examples. [2] Diabetes has a high impact on general populations, in particular in Western societies: The prevalence increases with age and reaches a level of about 10% at the age of 60 years and above in most populations. [1] Diabetes prevalence has been found to increase in all age groups during the last 20 years, and WHO has estimated that the number of diabetics in the world will more than double within the next 25 years from about 170 million in year 2000 to about 366 million in year 2030. [3] This increase of about 3.1% per year in prevalence has been claimed to fulfill the criteria for an epidemic, particularly of type 2 diabetes. [4,5] Others have claimed that an increase in prevalence alone cannot be taken as proof of an ongoing epidemic. [6] Also, prevention of diabetes (in particular type II) has attracted much attention with focus on the relation between lifestyle and risk of becoming a diabetic, and several countries have set up large scale prevention programs. From a medical point of view, the secondary prophylaxis is of high interest with focus on improving life quality and survival with diabetes through health initiatives ranging from motivational intervention, monitoring of disease, to new pharmaceutical treatments.

At the heart of these perspectives is the need to obtain valid and current estimates on trends in diabetes incidence, prevalence and mortality. Such estimates are, however, typically costly to obtain on a general population level, and are hence often missing. In this context, it has been suggested that pharmacoepidemiological databases could be used to study the dynamics of the diabetes epidemiology for large, well-defined populations, even when the databases only contain information on filled prescriptions. [7] A prime advantage of using pharmacoepidemiological databases for such epidemiological descriptions is obviously that these data are readily available. But the approach is not without its

challenges, since both date of onset as well as true, current disease status are not recorded in this type of pharmacoepidemiological databases. Misclassification is thus a potential problem, and the type of potential bias introduced by different strategies for determining treatment status at any given point in time must be carefully considered. Indeed, intuitively attractive approaches may introduce strong biases when studying calendar time trends, as discussed below.

Although we update and extend previously published results on the epidemiology of diabetes in the Funen population<sup>7</sup>, the primary objective of this paper is methodological. First, we investigate the impact of different choices of run-in periods, as this is a crucial methodological issue in many pharmacoepidemiological studies. Secondly, we study the relative contributions of current trends in incidence and mortality to the concurrent, observed rise in diabetes prevalence, ie. we focus on the question: Can prevalence rise while contemporary incidence and mortality rates remain constant?

## Methods

Odense Pharmaco-Epidemiologic Database (OPED) was used as data source. OPED contains information on all redemptions of subsidized and prescribed drugs at community pharmacies in the county of Fyn, Denmark [8] since 1992. It is a copy of the exhaustive electronic register maintained for administering all subsidies to pharmacies as well as prescribing physicians in the county. Information on the drug is entered directly into the electronic register when a prescription is processed and dispensed—which yields very high validity [9]—and in accordance with the Anatomical Therapeutic Chemical (ATC) classification system. [10] All anti-diabetic drugs are characterized by the first three characters being A10, insulin by A10A, oral anti-diabetics by A10B.

OPED contains information on birth date, gender, migration and death for all subjects living within the county in the period January 1, 1992 to December 31, 2003 in accordance with the Danish Central Person Register. All records allow unique tracking of individuals by means of the nationwide Civil Registration Number.

For each individual we determined presence and treatment status with respect to use of anti-diabetic medications on January 1 of each calendar year (the index date). We defined treatment status based on a previous run-in period's recordings of anti-diabetic drug dispensing, ie. if a subject had at least one prescription within the run-in period, the person was considered treated at the index date. Consequently, for each calendar year we excluded all who immigrated into Fyn during the run-in period, since we could not determine treatment status with certainty for immigrants without an observed redemption.

For those present at the index date and with a sufficient run-in period, we computed annual counts of these entities: prevalence ( $\mathbf{P}$ ), incidence ( $\mathbf{I}$ ), treatment cessation ( $\mathbf{C}$ ), deaths among all non-prevalents ( $\mathbf{D}_0$ ), deaths among prevalent ( $\mathbf{D}_1$ ), and deaths among incidents ( $\mathbf{D}_{\text{Inc}}$ ). An individual was considered prevalent on the index date if a redemption was observed in the associated run-in period. Incidence was defined as the number of subjects having a redemption in the year of interest, but not in the run-in period. Mortality counts were number of deaths in the year of interest stratified on prevalence status at the index date. Among those prevalent at the index date we defined treatment cessation as having been observed for a period which exceeded the length of the run-in period without any new redemptions and which overlapped the index date. In a sub-analysis of the annual transitions between treatment states (treated or not), we further subdivided the number of subjects who had discontinued treatment into two groups:

Those who re-initiated treatment within the calendar year, and those who did not.

For incidence and mortality among non-prevalents and prevalent, respectively, we computed rates based on counts and time at risk for the event in question. Time at risk was defined as time from index date to either event or censoring by migration, death, or end of year, whichever was first. Relative mortalities were obtained by comparing mortality rates among prevalent to rates among non-prevalents.

In the sub-analyses of insulin and oral anti-diabetic medications, only redemptions of the specific type were considered; for example, all redemptions of oral anti-diabetic agents were ignored when restricting the analysis to insulin.

Choosing the length of the run-in period is crucial in this study, as in most pharmacoepidemiological studies, since rates of exclusion must be balanced with misclassification. For the fixed run-in periods we did two separate analyses with lengths of one and two years, respectively. The length of one year allows comparison with previously published results [7], but we found it to be inadequate for accommodating a change in the Danish subsidy system taking place in 2000. We thus report on the effects of this change and provide estimates based on a two years run-in period, as these are unaffected by the switch. For illustrative purposes we also present estimates based on increasing the run-in period with calendar time, although these are invalid as they introduce time-dependent misclassification.

We finally investigated the extent to which the rise in prevalence could be explained by concurrent trends in incidence, mortality among treated and cessation rates, respectively. We did so by comparing the observed annual prevalences to projections based on fixing age-specific rates of incidence, mortality and cessation among treated to their values in 1994. Alternatively, incidence alone was fixed to its level in 1994 to study the isolated impact of a trend in incidence.

## Statistical analysis

Annual incidence, prevalence and mortality rates were determined within four age categories stratified by sex. Annual trend estimates was obtained from regression analysis with year and age categories (cut points at 15, 25, . . . , 85) as covariates stratified on gender. Logistic regression was employed for analyzing prevalence, Poisson regression for incidence and mortality rates. In the analyses of prevalence and incidence, correlations among individual's outcomes in subsequent years were allowed for by using robust standard errors. [11,12] For prevalence the trend was estimated as a linear trend on the log-odds-ratio scale, whereas trends in both incidence and mortality were estimated as linear trends on a log-rate-ratio scale. Cessation rates were not analyzed for trends as they are not of primary interest and in any case estimates would be based on small numbers. Estimates of age-adjusted annual trends with 95%-confidence intervals are given as odds-ratios and rate-ratios. All analyses were performed in Stata. 8 [13]

## Results

Basic characteristics of the studied population are presented in Tables 1(a+b) using a two year run-in period. Population size is almost constant with only a slight increase while the age-composition changes with an increase in the middle-aged group (40-64 years) and among children (<15), and a decline among younger adults (15-39 years). For both sexes, and in all age-groups, prevalence increases, most distinctly among middle-aged, as does incidence, with males aged 15-39 years as the only exception. The total number of deaths among untreated across all age groups declines over the study period, in particular among males above age 65. Mortality among treated increases in absolute numbers, but declines relatively when compared to the concurrent increase in prevalence. A graphical presentation of relative measures is given in Figure 1.

[Tables 1(a+b) about here]

[Figure 1 about here]

Careful inspection of Tables 1(a+b) reveals that bookkeeping with respect to prevalence does not tally in the following sense: The prevalence of a given year does not correspond to the prevalence plus incidence minus deaths among treated of the previous year. What is missing is the count of apparent treatment cessations. In Table 2 the bookkeeping is carried out including the annual number of treatment cessations and correcting the number of deaths among the prevalents accordingly. The final column of Table 2 shows that the discrepancy between the predicted prevalence according to this bookkeeping principle and the

actually recorded prevalence is small. The difference is due to censoring induced by migration, since the cohort is open. .

[Table 2 about here]

Table 3 gives estimates of age- and sex-adjusted trends. Both prevalence and incidence trend estimates are positive and statistically significant, regardless of gender, anti-diabetic drug type, and length of run-in. Mortality declines among treated (except for males with respect to insulin when using a two year run-in period) over the study period, and notably also relative to the non-treated population.

[Table 3 about here]

### **Methodological considerations I: length of run-in**

Table 4 shows the annual count of patients defined as having discontinued treatment based on a gap of either one year or two years between last redemption in the run-in period and first redemption in the following year. Using a gap of one year leads to a peak in 2001 reflecting a change in reimbursement policy in 2000, so apparently some have stockpiled in early 2000 followed by a longer period without redemptions. This effect is smoothed out by using a gap of two years, i.e. the two year run-in period defines treatment status without obvious time-dependent misclassification.

[Table 4 about here]

As seen from Table 3 there is, however, virtually no difference in trend estimates of prevalence, incidence and mortality, whether we choose a one or two year year long run-in period. Only exception is for incidence, and this is a result of a high incidence in 1993. With a two year run-in period, we are forced to omit this year when estimating trends, and hence the trend is attenuated.

If, however, we were to naively use all available information prior to a given index date to determine treatment status, trend estimates of prevalence become markedly increased, and trend estimates for incidence decreased. The relative biases of this naive approach range from 20%-40% for these parameters. Trends in mortality are unaffected by using a variable length run-in period.

### **Methodological considerations II: Anatomy of an “epidemic”**

Above we observed an increase in prevalence and incidence accompanied by a decrease in mortality among treated. A common interpretation would be to explain the trend in prevalence with the concurrent trends

in incidence and mortality. This would however be an oversimplification if prevalence could rise while incidence and mortality rates remained constant. To examine whether this actually occurred during the study period, we compared the observed number of prevalents over the study period to the number of prevalents that would be predicted if incidence, mortality among treated, and cessation rates all had remained constant at their level in 1994. Alternatively, we only fixed incidence rate at its 1994-level, while allowing mortality among treated and cessation rates to vary as observed. The resulting projections are shown in Figure 2.

[Figure 2 about here]

For the period 1995-1999, agreement is very good between observed and projected prevalence among females, while poorer among males. For both genders the discrepancy increases from year 2000. Keeping only incidence fixed improves agreement between observed and projected prevalence substantially, in particular among men, ie. for men the combined decline in mortality and cessation explain more of the observed rise in prevalence than for women. The bulk of the rise in prevalence is thus not due to contemporary changes in incidence or mortality rates, but to incident subjects outnumbering deaths among treated and treatment cessations.

## Discussion

Throughout the study period, prevalence of use of anti-diabetic agents increased both in absolute and relative terms, as well as within and across age strata. The corresponding incidence also increased, while mortality among treated declined, also relative to mortality among untreated. This agrees with previous results [7], except with respect to incidence.

The data used are generally of high validity, as they result from automated registration of economic transaction data; near complete coverage, except for medications filled at a pharmacy outside the subjects residential county, which for all drugs was found to not exceed 0.5% [9]; the information on migration and vital statistics is identical to the relevant subset of the registration at Statistics Denmark.

The major limitation of the present study is its reliance on drug redemptions. Obtaining a drug is not equivalent to using it, and when used there is no observation of end of use. Consequently, we investigated the impact of different treatment classification rules and found that trend estimates for prevalence, incidence, and mortality were largely unaffected, except for the flawed principle of using a run-in period of varying length.

Further, our estimates for prevalence, incidence, and mortality by definition only concern subjects with treated diabetes, i.e. our results do not apply to undiagnosed diabetics, nor to diabetics treated exclusively with diet. These subjects are misclassified in the sense that they are considered part of the general, non-diabetic population in the present study. Still, the prevalence level found here is in line with other contemporary studies in Denmark [14, 15], and we believe our results to be of interest, as they pertain to that group of diabetics which must generally be considered most severely afflicted.

A third consequence of relying on a pharmacoepidemiological database is the inability to control for date of diagnosis in the analysis. In particular, in the analysis of mortality among treated we could not account for duration of disease, and part of the explanation for the apparently decreasing mortality could be earlier treatment of patients with less severe diabetes.

The trend in incidence identified in this study was comparable to the one found for the period 1993-1999, reported in [7], although slightly attenuated. The attenuation occurred both because the year 1993 was excluded in the present study using a two year run-in period, and because incidence began a steeper increase in 2000. Combined with the enlarged sample size due to a longer study period, this made the trend statistically significant. The increase in incidence observed after year 2000 could hallmark the onset of a true diabetes epidemic, possibly reflecting an impact from a sustained obesity epidemic on onset of diabetes or other changes in lifestyle—although postulating a general obesity epidemic in Denmark would also appear to be an over-simplification according to a recent study showing the existence of large heterogeneity in obesity trends

cohorts. [16] That an epidemic of diabetes could be looming was supported by the finding that fixing incidence at its level in 1994 could predict prevalence with good precision until 2000, at which point the observed prevalence started to outrun projected prevalence. It may, however, be speculated that new diagnostic criteria, which eased detection of new cases and was implemented in 1999 in Denmark, could have contributed to the observed increase in incidence. If this is the case, it should be expected that the rise in incidence levels off again, insofar as it is due to the new diagnostic criteria. Note, however, that our considerations about the relation of incidence, prevalence and mortality are independent of any assumptions about the causal reason for a shift in incidence, be it a real increase in subjects developing diabetes, a change in diagnostic procedures, criteria or habits, or a change in treatment practices.

Whatever the explanation, the trend in incidence should not be considered the primary reason for the concurrent increase in prevalence. Instead, the primary explanation is the annual surplus of subjects initiating treatment for diabetes compared to those leaving treatment which must be considered a legacy of

past diabetes incidence and treatment. The surplus may be expected to increase further, if current trends in incidence and mortality persist, and projections by WHO and others may thus severely underestimate the true magnitude of diabetes prevalence ten to twenty years from now, built as they are on stationary age-specific incidence and mortality estimates. [3]

In conclusion, we suggest that pharmacoepidemiologic databases can provide useful information on current trends in diabetes epidemiology if analyzed properly, and at a much lower cost than ordinary studies. The major drawback is the lack of insight in the underlying reasons for any observed trends—an insight which can only be obtained on a regular basis if a population-wide case database collecting information on all incident and prevalent diabetes cases is established.

### **Competing interests**

The authors declare that they have no competing interests.

### **Authors' contributions**

Støvring had the original idea for the study, carried out all analyses and drafted the original and revised manuscripts. Andersen provided access to data. The planning of analyses and interpretation of the data was the product of discussions involving all authors, all of whom also commented on drafts of the paper. All authors have seen and approved the final version.

### **References**

1. Harris M, Hadden W, Knowler W, Bennet P: **Prevalence of diabetes and impaired glucose tolerance and plasma glucose levels in U.S. population aged 20-74 yr.** *Diabetes* 1987, **36**:523–34.
2. Stratton I, Adler A, Neil H, Matthews D, Manley S, Cull C, Hadden D, Turner R, Holman R: **Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study.** *British Medical Journal* 2000, **321**(7258):405–12.
3. Wild S, Roglic G, Green A, Sicree R, King H: **Global prevalence of diabetes: estimates for the year 2000 and projections for 2030.** *Diabetes Care* 2004, **27**(5):1047–53.

4. Mokdad A, Bowman B, Ford E, Vinicor F, Marks J, Koplan J: **The continuing epidemics of obesity and diabetes in the United States.** *JAMA* 2001, **286**:1195–200.
5. Colagiuri S, Borch-Johnsen K, Glümer C, Vistisen D: **There really is an epidemic of type 2 diabetes.** *Diabetologia* 2005, [<http://dx.doi.org/10.1007/s00125-005-1843-y>].
6. Green A, Støvring H, Andersen M, Beck-Nielsen H: **The epidemic of type 2 diabetes is a statistical artefact.** *Diabetologia* 2005, [<http://dx.doi.org/10.1007/s00125-005-1841-0>].
7. Støvring H, Andersen M, Beck-Nielsen H, Green A, Vach W: **Rising prevalence of diabetes: evidence from a Danish pharmaco-epidemiological database.** *Lancet* 2003, **362**(9383):537–8.
8. Hallas J, Nissen A: **Individualized drug utilization statistics. Analysing a population's drug use from the perspective of individual users.** *Eur J Clin Pharmacol* 1994, **47**(4):367–372.
9. Gaist D, Sorensen H, Hallas J: **The Danish prescription registries.** *Dan Med Bull* 1997, **44**(4):445–8, [<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?cmd=%prlinks&dbfrom=pubmed&retmode=ref&id=9377907>].
10. The WHO Collaborating Centre for Drug Statistics Methodology: *ATC index with DDDs and Guidelines for ATC classification and DDD assignment.* Oslo 2001.
11. Huber PJ: **The behavior of maximum likelihood estimates under nonstandard conditions.** In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics*, Berkeley, Calif.: Univ. California Press 1967:221–233.
12. White H: **Maximum likelihood estimation of misspecified models.** *Econometrica* 1982, **50**:1–25.
13. StataCorp: *Stata Statistical Software: Release 8.0.* College Station, Texas: Stata Corporation 2003.
14. Drivsholm TB, Frederiksen K, de Fine Olivarius N, Ødegaard B, Kristensen JK: **[The prevalence of diabetes in Denmark. Development of a method for a registry-based assessment].** *Ugeskr Laeger* 2003, **165**(29):2887–2891.
15. Danish Medicines Agency: **Antidiabetika: Forbruget af antidiabetika 1994 - 1997 (Antidiabetics: The use of antidiabetics 1994 - 1997).** Brønshøj, Denmark 1998. [(In Danish)].

16. Heitmann B, Stroger U, Mikkelsen K, Holst C, Sorensen T: **Large heterogeneity of the obesity epidemic in Danish adults.** *Public Health Nutr* 2004, **7**(3):453–60, [<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?cmd=%prlinks\&dbfrom=pubmed\&retmode=ref\&id=15153276>].

## **Figures**

### **Figure 1**

Prevalence, incidence and mortality estimates stratified by gender and age, Fyn County, Denmark, 1994-2003. Line-styles indicates age-categories in years: — 0-14, - - 15-39, ··· 40-64, - · - 65+.

### **Figure 2**

Observed and predicted prevalences with respect to use of anti-diabetic drugs, Fyn County, Denmark, 1994-2003. Fixation of rates is with respect to their 1994-level.

## **Tables**

### **Table 1(a) - Characteristics of source population**

Size of population, prevalence, incidence, and mortality with respect to use of anti-diabetic drugs, Fyn County, Denmark, 1994-2003.

### **Table 1(b) - Characteristics of source population (totals)**

Total size of population, prevalence, incidence, and mortality with respect to use of anti-diabetic drugs, Fyn County, Denmark, 1994-2003.

### **Table 2 - Bookkeeping for size of total prevalent population**

Bookkeeping for size of total prevalent population with respect to anti-diabetic treatment, Fyn County, Denmark, 1994-2003.

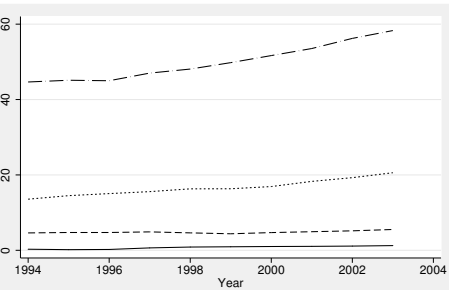
### **Table 3 - Trend estimates for incidence, prevalence, and mortality with respect to treatment with anti-diabetic agents**

Estimated age-adjusted, annual trends with 95%-confidence intervals for different categories of anti-diabetic drugs and with different length of run-in periods, Fyn County, Denmark, 1992-2003.

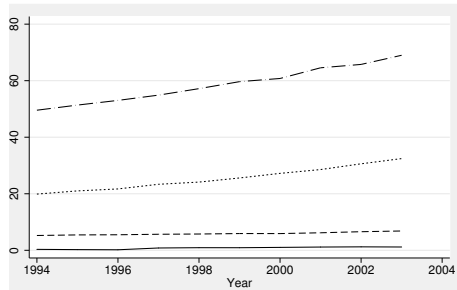
### **Table 4 - Discontinuation of anti-diabetic treatment**

Number of patients discontinuing anti-diabetic treatment, Fyn County, Denmark, 1994-2003.

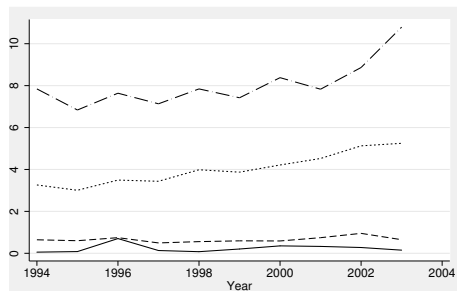
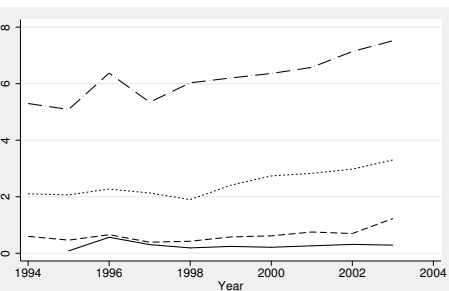
Females



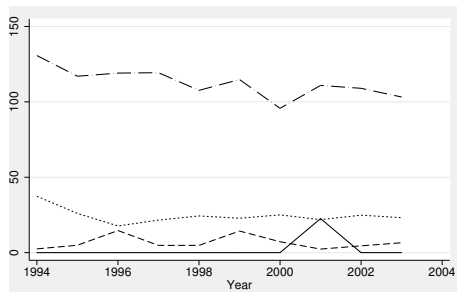
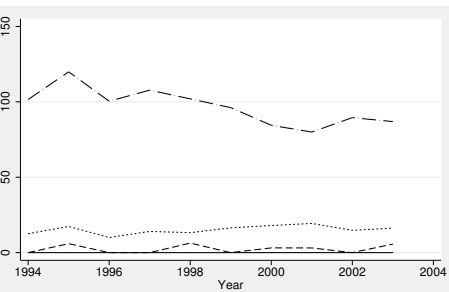
Males



(a) Prevalence per 1,000 persons

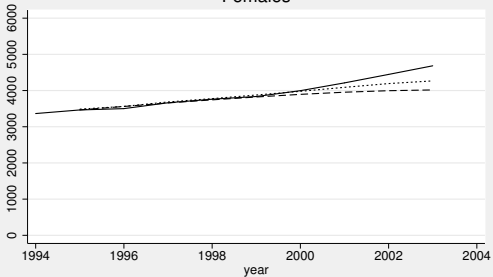


(b) Incidence per 1,000 person-years

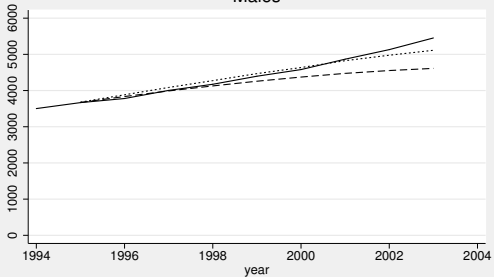


(c) Mortality among treated per 1,000 person-years

### Females



### Males



— Observed prevalence      - - - - - All rates fixed  
..... Only incidence rate fixed

— Observed prevalence      - - - - - All rates fixed  
..... Only incidence rate fixed

Figure 2

**Additional files provided with this submission:**

Additional file 4 : table4.pdf : 11Kb

<http://www.pophealthmetrics.com/imedia/1627233618116971/sup4.PDF>

Additional file 3 : table3.pdf : 11Kb

<http://www.pophealthmetrics.com/imedia/4298561811169710/sup3.PDF>

Additional file 2 : table2.pdf : 18Kb

<http://www.pophealthmetrics.com/imedia/1899358798116971/sup2.PDF>

Additional file 1 : table1.pdf : 26Kb

<http://www.pophealthmetrics.com/imedia/8056711751169710/sup1.PDF>