

IFEDH AP8: Influenza

Florian Endel

13. Mai 2012

Inhaltsverzeichnis

1 Besprechung 11.05.2012 (HVB)	1
2 Diagnosen von Akuterkrankungen	2
2.1 ATC-ICD 'akut'	2
2.2 Vorgehen	2
2.3 Methodik	2
2.4 erste Analyse	3
2.4.1 nicht akute Erkrankungen	5

1 Besprechung 11.05.2012 (HVB)

Bei der Besprechung wurde das Influenza Projekt im Kontext eines größeren und umfangreicheren Influenza Projekts vorgestellt. Vor allem die Probleme aufgrund fehlender Daten und unsicherem bzw. ungenauem Wissens zu Influenza wurden diskutiert. Details die nicht die Datenerhebung betreffen werden hier nicht weiter beschrieben.

Folgende Daten wurden bisher verwendet:

- Wetter (hohe Werte)
- Krankenstände (GAP-DRG)
- Sentinel Österreich ([LINK](#))
- MA15: Hochrechnung von wöchentlichen Fällen von Influenza und ILI für Wien
- Expertenschätzungen (5% erkranken pro Jahr)
- Literatur (z.B. Kinder sind 6x häufiger betroffen)

Keine der genannten Quellen konnte ein ausreichend detailliertes Bild der Influenzaerkrankungen in Österreich darstellen. Auch ein Zusammenhang mit dem Wetter (Temperatur und Luftfeuchtigkeit) konnte nicht festgestellt werden.

Aus der Sentinel Datensammlung sind allerdings die Zeiträume (Wochen) der Grippewellen bekannt. Ein Vergleich dieser Zeiträume mit der Anzahl an Krankenständen im selben bzw. den umgebenden Zeiträumen lässt einen Zusammenhang vermuten. Ein ähnlicher Vergleich soll daher für Diagnosen von Akuterkrankungen erstellt werden (2).

2 Diagnosen von Akuterkrankungen

Das Ziel ist die Anzahl der Personen, die im Laufe einer Grippewelle (Dauer 19 Wochen) an 'echter' Grippe erkranken, abzuschätzen. Da der Zeitraum der Grippewelle aus Sentinel¹ bekannt ist und Diagnosen für Akuterkrankungen basierend auf ATC-ICD[1] in GAP-DRG vorhanden sind, wurde vorgeschlagen aus diesen Zusammenhängen die Anzahl zu schätzen.

2.1 ATC-ICD 'akut'

Für die Sozialversicherungsträger NoeGKK, KGKK, SVB, SVA (12, 16, 40, 50) liegt eine ATC-ICD Implementation² mit wöchentlichen Diagnosen vor. Aufgrund von Datenqualitätsproblemen bei den Datumsangaben der anderen Träger konnten diese genauen Angaben nur in dieser Auswahl umgesetzt werden. Wie in der entsprechenden Analyse der Datumsverteilungen³ nachzulesen ist, wurden nur Rezeptdaten und Leistungsdaten im niedergelassenen Bereich beachtet, nicht aber Krankenstand- und Krankenhausdaten.

In der ATC-ICD Implementation der Akuterkrankungen liegen daher Diagnosen (ICD9 3-steller) pro Woche und Person vor. Da die, im ATC-ICD Projekt erzeugten Zuordnungstabellen für gewisse Alterskohorten und Geschlechtsgruppen erstellt wurden, können starke Unterschiede bei Personen unterschiedlicher Altersgruppen auftreten.

Es ist daher jedenfalls notwendig die Zuordnungen, Zuordnungswahrscheinlichkeiten und Quelle (ATC-Codes, Medikamente) pro Kohorte zu analysieren um systematische Fehler, die durch diese Form der Datenaufbereitung entstanden sind, zu vermeiden (vgl. Dokument von Alexander Eisl).

2.2 Vorgehen

Das Arbeiten mit den Daten beginnt sofort nach der Besprechung. D.h. das Ausformulieren der HTA Fragestellung und etwaige andere Vorarbeiten werden parallel dazu ausgeführt. Dies ist notwendig, um in der vorhandenen Zeit an Ergebnisse zu kommen. Allgemein ist anzumerken, dass nur aufgrund der fertigen Datenbasis GAP-DRG und den umfangreichen Vorarbeiten (z.B. Bereinigungen, Record Linkage, ATC-ICD, Integration von ATC-ICD, DQA, etc.) und vorhandenen Erfahrungen in dieser Zeit mit Ergebnissen zu rechnen ist.

Obwohl auf die erwähnten notwendigen Analysen der Daten und deren Herkunft nicht verzichtet werden sollte, werden im ersten Schritt einfache Datensätze abgefragt und analysiert um die allgemeine Machbarkeit zu überprüfen. Sollten die erwarteten Strukturen nicht zu finden sein, sind umfangreiche Qualitätsanalysen (für dieses Projekt) auch nicht mehr notwendig. Falls aber Informationen aus diesem Datensatz verwendet werden, sind die entsprechenden Angaben notwendig.

2.3 Methodik

Ausgehend von der Anzahl von Personen mit Akuterkrankungen pro Woche wird versucht ein Unterschied zwischen Wochen in und außerhalb einer Grippewelle zu erkennen. Es wird angenommen, dass die meisten der Personen die in diesen Zeiträumen zusätzliche Akuterkrankungen erleiden tatsächlich auch an Grippe erkrankt sind. Ökologische Fehlschlüsse und saisonale Schwankungen die unabhängig von Grippewellen auftreten werden dabei (vorerst) ignoriert.

¹Referenz

²ATC-ICD *neu*: http://gapdrg.endel.at/dokuwiki/doku.php/gapdrg:projekte:atc-icd_neu

³Analyse der Datumsverteilungen: http://gapdrg.endel.at/dokuwiki/doku.php/gapdrg:projekte:dqa#analyse_der_leistungsdaten_pro_traeger

2.4 erste Analyse

Es werden zuerst die notwendigen Daten erfasst:

```
1 — Anzahl der Personen mit Akuterkrankungen pro Woche
2 SELECT diag_year, diag_week, date, count(DISTINCT pers_id) AS anz
3 FROM gapdrg.atcid_akut_date
4 WHERE icd9_int BETWEEN 460 AND 466
5 OR icd9_int BETWEEN 480 AND 487
6 GROUP BY date, diag_year, diag_week;
```

Listing 1: Verteilung Akuterkrankungen

Die Daten werden in R importiert und ein Überblick dargestellt:

```
dat = read.csv('./data/01 akut.csv')
names(dat) = c('Jahr', 'Woche', 'Datum', 'Diagnosen')

dat$Datum = as.Date(as.character(dat$Datum), format='%Y-%m-%d')

str(dat, vec.len=2)

## 'data.frame': 104 obs. of 4 variables:
## $ Jahr      : int  2006 2006 2006 2006 2006 ...
## $ Woche     : int   1  2  3  4  5 ...
## $ Datum     : Date, format: "2006-01-01" "2006-01-08" ...
## $ Diagnosen: int  32518 50781 45476 44333 48584 ...

summary(dat)

##      Jahr      Woche      Datum      Diagnosen
## Min.   :2006   Min.    : 1.0   Min.   :2006-01-01   Min.   :21657
## 1st Qu.:2006   1st Qu.:13.8   1st Qu.:2006-06-30   1st Qu.:32905
## Median :2006   Median :26.5   Median :2006-12-28   Median :38516
## Mean   :2006   Mean    :26.5   Mean   :2006-12-28   Mean   :38851
## 3rd Qu.:2007   3rd Qu.:39.2   3rd Qu.:2007-06-26   3rd Qu.:44587
## Max.   :2007   Max.    :52.0   Max.   :2007-12-24   Max.   :59776
```

Der Verlauf für die Jahre 2006 und 2007 mit loess smoothing:

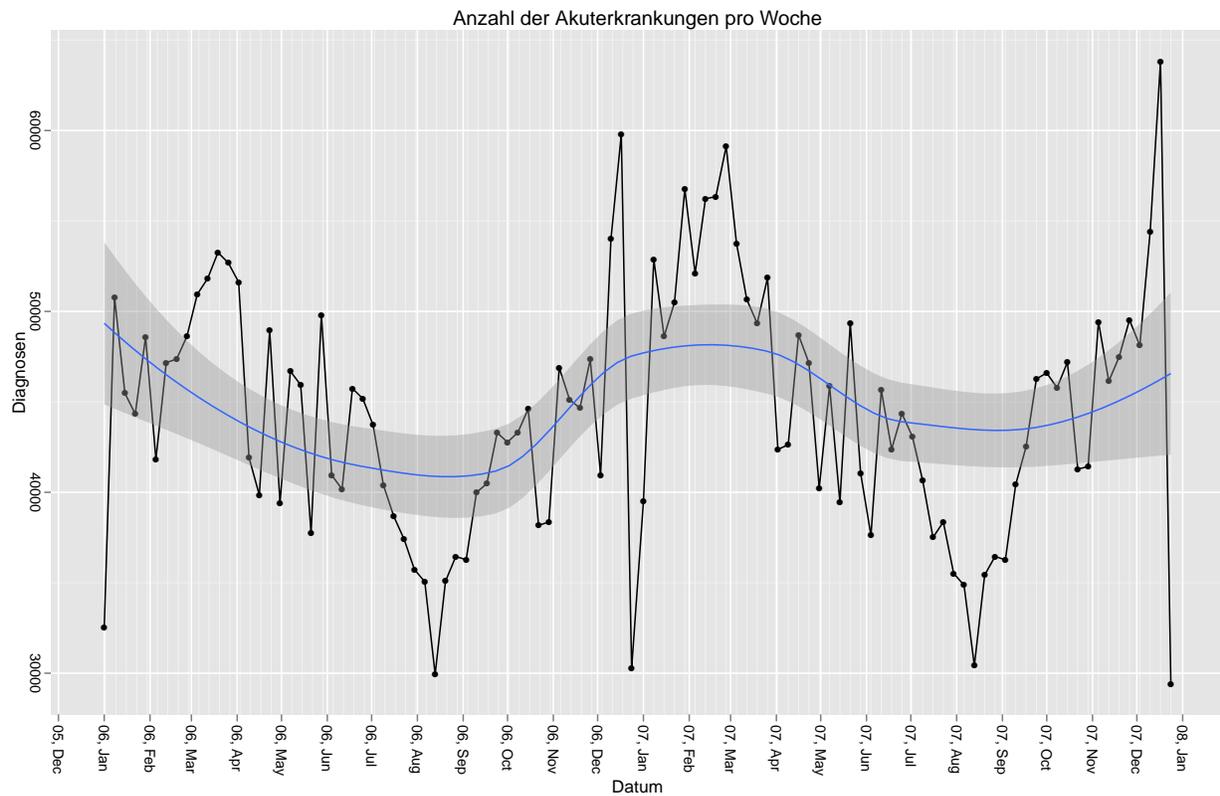
```
p = ggplot(dat, aes(x=Datum, y=Diagnosen))
p = p + opts(title='Anzahl der Akuterkrankungen pro Woche')
p = p + geom_line()
p = p + geom_point()
p = p + geom_smooth(method='loess')
p = p + scale_x_date(
  breaks = date_breaks("month"),
  minor_breaks = date_breaks("weeks"),
  labels = date_format("%y, %b")
)
```

```

p = p + opts(
  axis.text.x=theme_text(angle=-90, hjust=0),
  axis.text.y=theme_text(angle=-90, hjust=0.5)
)

# beide Jahre
print(p)

```



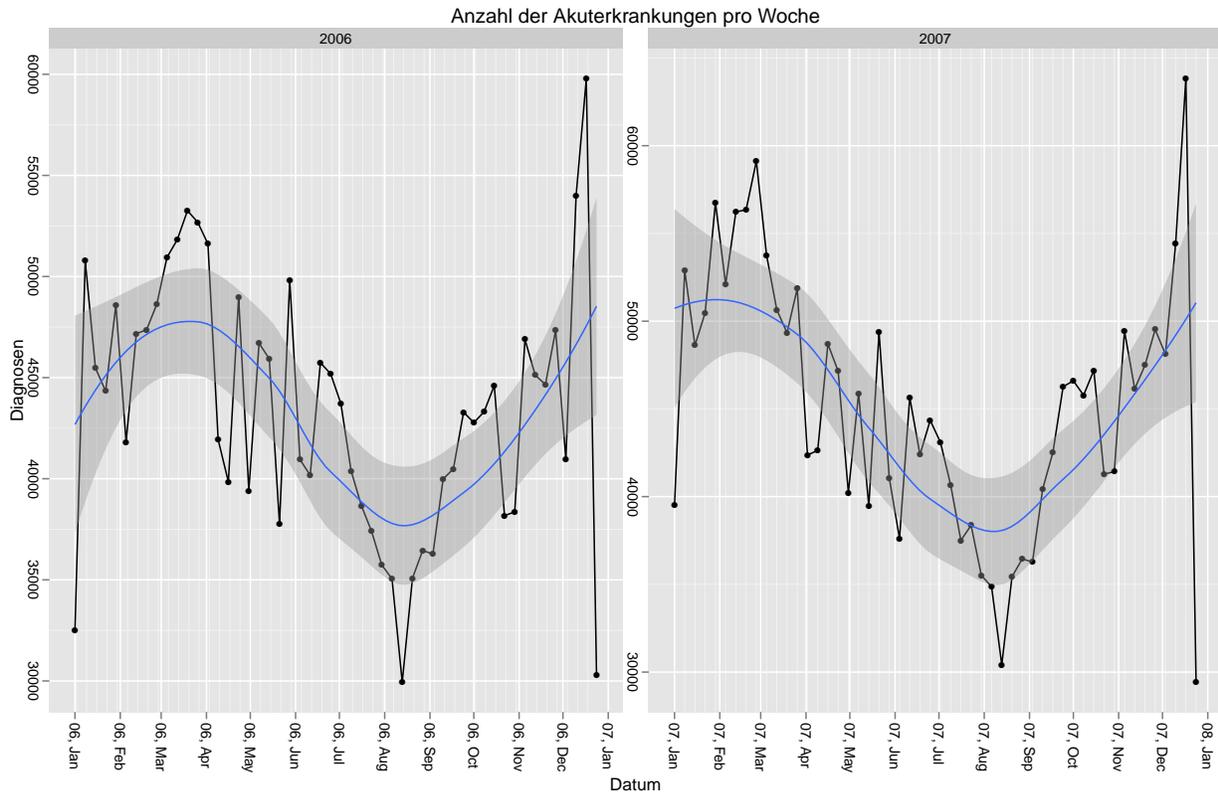
Im Hintergrundraster sind Monate und Wochen eingezeichnet.

Da es anscheinend einen Niveauunterschied zwischen den beiden Jahren gibt, werden diese unabhängig mit angepasster Skalierung dargestellt.

```

# getrennt und rescale
p = p + facet_wrap(~Jahr, scale = c("free"))
print(p)

```



Gewisse systematische Abweichungen „nach oben“ ist in beiden Jahren im Bereich Februar/März zu sehen. Sehr auffällig sind massive Ausreißer Mitte August und Ende Dezember. Möglicherweise kann ein Zusammenhang mit beschrieben werden, sobald die Perioden (Wochen) der Grippewellen bekannt sind.

2.4.1 nicht akute Erkrankungen

In diesem Abschnitt werden alle Diagnosen, außer die akuten (potentiellen) Influenzaerkrankungen erfasst und dargestellt. Sollten die Grippewellen tatsächlich einen Einfluss auf die absolute Anzahl an Diagnosen haben, müsste der Verlauf der anderen Erkrankungen weniger starke Schwankungen nach oben zeigen. Sollte dies nicht der Fall sein, kann eher von einer jahreszeitabhängigen Variation, als einem Einfluss der Grippewelle ausgegangen werden.

```

1 — Anzahl der Personen mit Akuterkrankungen pro Woche
2 SELECT diag_year, diag_week, date, count(DISTINCT pers_id) AS anz
3 FROM gapdrg.atcid_akut_date
4 WHERE icd9_int NOT BETWEEN 460 AND 466
5 OR icd9_int NOT BETWEEN 480 AND 487
6 GROUP BY date, diag_year, diag_week;

```

Listing 2: Verteilung restlicher Diagnosen

Die Daten werden in R importiert und ein Überblick dargestellt:

```

dat.rest = read.csv('./data/02 nicht akut.csv')
names(dat.rest) = c('Jahr', 'Woche', 'Datum', 'Diagnosen')

```

```

dat.rest$Datum = as.Date(as.character(dat.rest$Datum), format='%Y-%m-%d')

str(dat.rest, vec.len=2)

## 'data.frame': 104 obs. of 4 variables:
## $ Jahr      : int  2006 2006 2006 2006 2006 ...
## $ Woche     : int  1 2 3 4 5 ...
## $ Datum     : Date, format: "2006-01-01" "2006-01-08" ...
## $ Diagnosen: int  146902 259683 230670 220389 240006 ...

summary(dat.rest)

##      Jahr      Woche      Datum      Diagnosen
## Min.   :2006   Min.    : 1.0   Min.   :2006-01-01   Min.   :115121
## 1st Qu.:2006   1st Qu.:13.8   1st Qu.:2006-06-30   1st Qu.:210349
## Median :2006   Median :26.5   Median :2006-12-28   Median :229342
## Mean   :2006   Mean    :26.5   Mean    :2006-12-28   Mean    :226402
## 3rd Qu.:2007   3rd Qu.:39.2   3rd Qu.:2007-06-26   3rd Qu.:240267
## Max.   :2007   Max.    :52.0   Max.    :2007-12-24   Max.    :316183

```

Der Verlauf für die Jahre 2006 und 2007 mit loess smoothing:

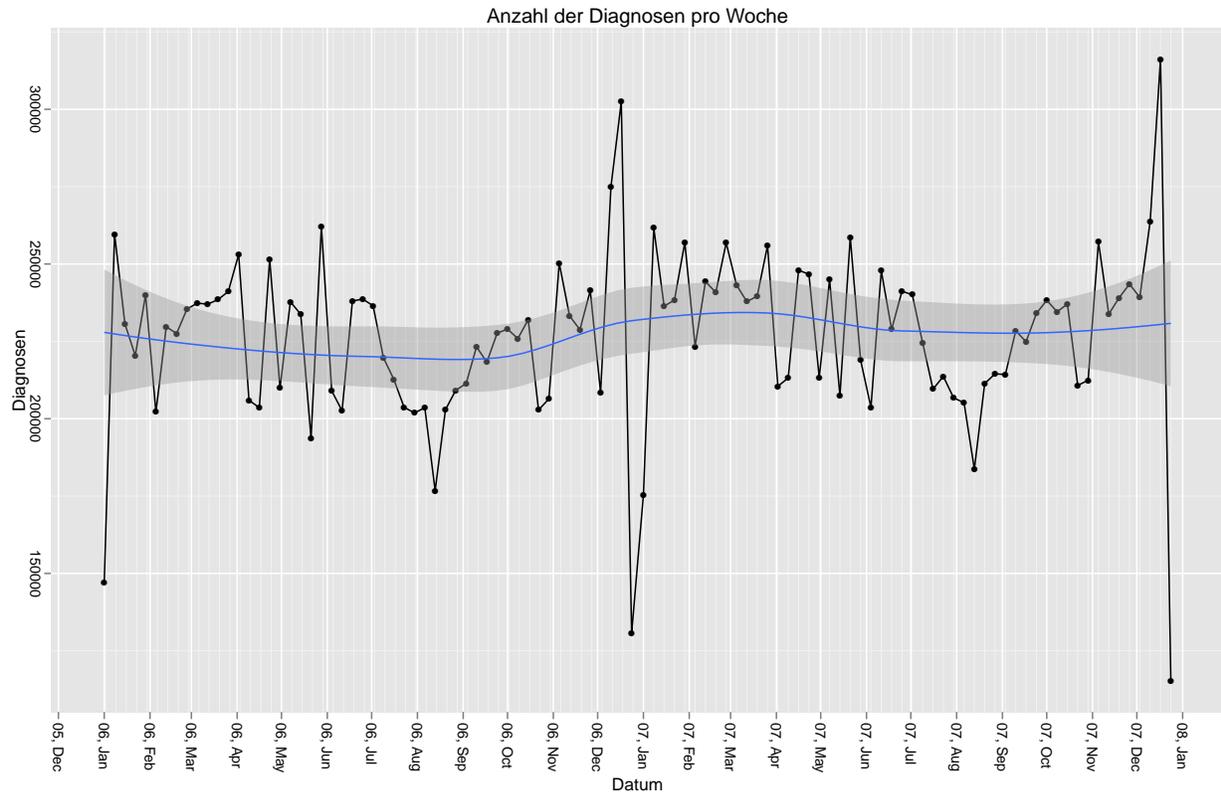
```

p = ggplot(dat.rest, aes(x=Datum, y=Diagnosen))
p = p + opts(title='Anzahl der Diagnosen pro Woche')
p = p + geom_line()
p = p + geom_point()
p = p + geom_smooth(method='loess')
p = p + scale_x_date(
  breaks = date_breaks("month"),
  minor_breaks = date_breaks("weeks"),
  labels = date_format("%y, %b")
)

p = p + opts(
  axis.text.x=theme_text(angle=-90, hjust=0),
  axis.text.y=theme_text(angle=-90, hjust=0.5)
)

# beide Jahre
print(p)

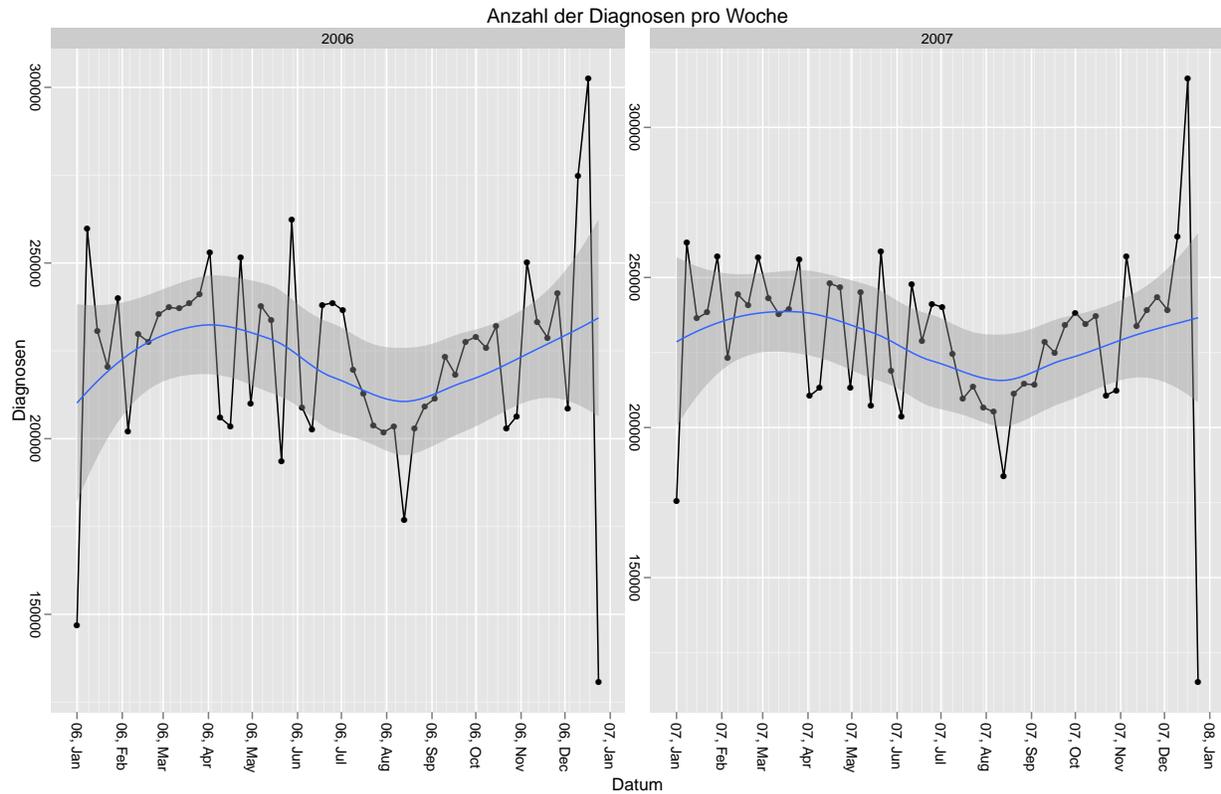
```



Im Hintergrundraster sind Monate und Wochen eingezeichnet.

Da es anscheinend einen Niveauunterschied zwischen den beiden Jahren gibt, werden diese unabhängig mit angepasster Skalierung dargestellt.

```
# getrennt und rescale
p = p + facet_wrap(~Jahr, scale = c("free"))
print(p)
```



Es sind wieder die massiven Ausreißer in einigen Wochen zu sehen. Allgemein ist zu erkennen, dass die jahreszeitlichen Schwankungen geringer ausfallen und die (systematischen?) Abweichungen im Februar/März nicht zu erkennen sind.

Literatur

- [1] Filzmoser et al. Atc → icd: Determination of the reliability for predicting the icd code from the atc code. 12 2009. Available from: http://www.hauptverband.at/portal27/portal/hvbportal/channel_content/cmsWindow?action=2&p_menuid=72727&p_tabid=5.