Incidence and Prevalence Estimations Based on Claims Data – New Methodological Considerations

Prosser C[1], Altevers J[1], Hickstein L[2,3], Walker J[2,3], Braun S[1]
[1]Xcenda GmbH, Hannover, Germany, [2]Elsevier Health Analytics, Berlin, Germany, [3]Health Risk Institute, Berlin, Germany

BACKGROUND

- Scientific analyses with claims data such as burden of disease analyses are often based on incidence and prevalence estimates.
- Prevalent cases in claims databases are typically identified by diagnoses, medications, and/or procedures in a specific observation period (eg, 1 calendar year).
- Incident cases, especially in chronic conditions, are usually derived by excluding patients already diagnosed or treated in a certain pre-observation period from the sample of prevalent cases.
- For incidence estimations in chronic diseases, latest methodological considerations[1,2] indicate that the pre-observation period should be extended as much as possible to prevent overestimations.
- So far, such methodological considerations have not been adopted for prevalence estimations in chronic diseases. By extending the observation period, underestimations of prevalence may be prevented, as it is assumed that chronic diseases identified in previous years persist until the prevalence year (**Figure 1**).

Figure 1. Classification of individuals as prevalent, incident, or healthy depending on a diagnosis in the prevalence year and/or pre-observation period for chronic conditions

		Diagnosis in prevalence year	
		Yes	No
Diagnosis in pre-observation period	Yes	prevalent	prevalent
	No	prevalent and incident	healthy

- As highlighted, individuals diagnosed with a chronic condition in the pre-observation period but not in the prevalence year are still considered to be prevalent in the prevalence year. Therefore, the observation period for the prevalence estimation should not be restricted to the prevalence year but should also include a pre-observation period, especially if both the incidence and prevalence are estimated.

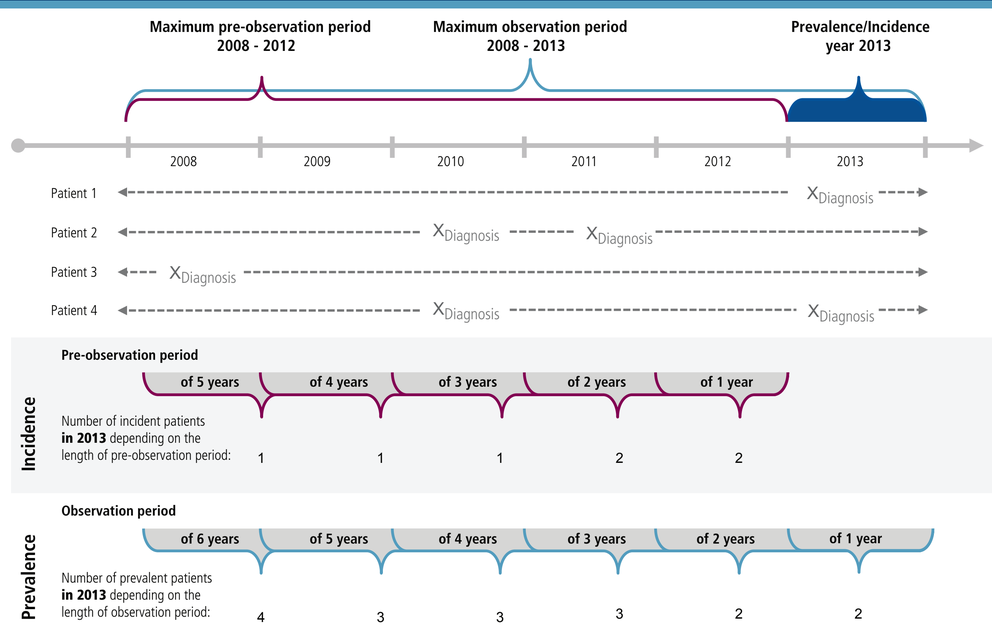
OBJECTIVE

- The aim of this methodological analysis was to evaluate the impact of expanding the observation period for the incidence as well as the prevalence estimates.

METHODS

- The analysis was conducted using the Health Risk Institute (HRI) Research Database which comprises claims data from 75 different German health insurances on an anonymized individual level. The analysis sample is representative for the German population in terms of age and gender (structure of age and gender as per 31 December 2013 / DESTATIS), and the insured individuals are distributed all over Germany.
- Individuals continuously insured and alive from 2008 until 2013 were included (n=3,026,154).
- The analysis focused on the chronic diseases diabetes mellitus (DM) and multiple sclerosis (MS) in 2013 in Germany. The ICD-10-GM codes E10-E14 were used to identify DM patients and G35 for MS patients in the inpatient (main or secondary diagnosis) or outpatient (verified diagnosis) sector.
- Individuals were considered to be prevalent in 2013 if they had at least one of these ICD-10-GM codes in the observation period.
- Individuals were defined as incident in 2013 if they had at least one of these ICD-10-GM codes in 2013 but not in the pre-observation period.
- Changes in incidence due to different lengths of pre-observation periods (1 to 5 years) before diagnosis in 2013 were assessed. Correspondingly, the prevalence estimation for 2013 was varied by expanding the observation period from exclusively 2013 stepwise to the whole available pre-observation period of 2008-2013 (**Figure 2**).
- Additionally, the impact on the proportion of incident cases within the prevalence was assessed.
- To validate the DM and MS diagnoses, a sensitivity analysis using a stricter case definition was performed. Diagnoses in the outpatient sector (verified diagnosis) had to be confirmed by at least one additional outpatient diagnosis or one prescription of DM/MS-specific medications[3] in the same or following 3 quarters.

Figure 2. Classification of patients as incident and/or prevalent depending on the length of (pre-) observation period

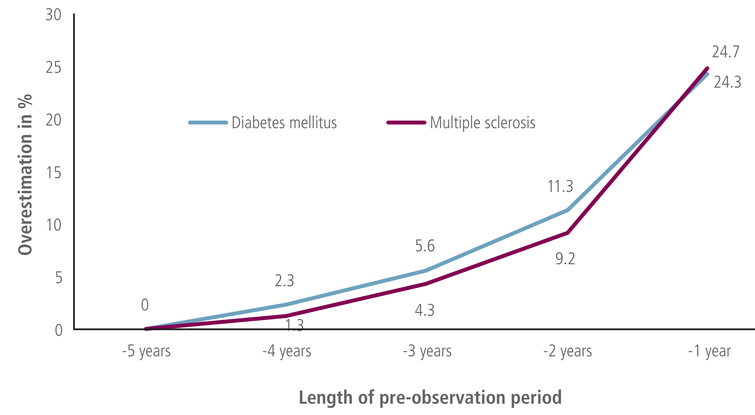


RESULTS

Change in Incidence

- Using a 5-year pre-observation period as the gold standard, the DM incidence was overestimated by 24% when a 1-year pre-observation period was applied. The extent of overestimation decreased steadily and approximately halved with each additional year of pre-observation.
- A similar magnitude and trend of overestimation was observed in MS. The overestimation ranged from 25% when using a 1-year pre-observation period to 1% when using a 4-year pre-observation period (**Figure 3**).
- Thus, with each year added to the pre-observation period, additional misclassifications of patients as incident were revealed.

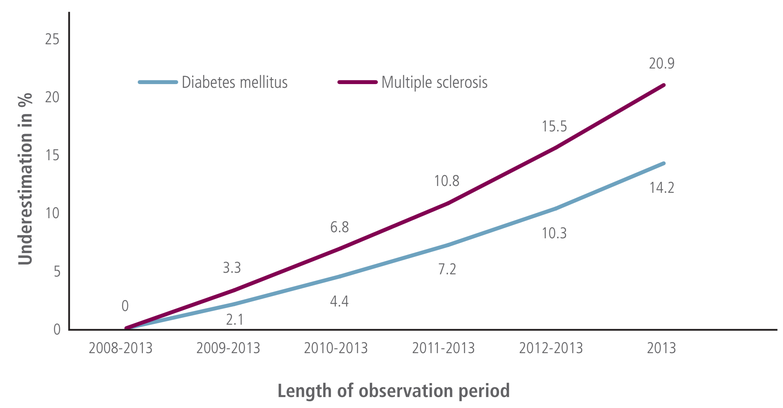
Figure 3. Overestimation of incidence estimates depending on different pre-observation periods for DM and MS in 2013 in Germany (reference=5-year pre-observation period)



Change in Prevalence

- Using the number of prevalent patients identified from 2008 to 2013 as the gold standard, the DM prevalence was underestimated by 14% when the observation period was restricted exclusively to the year 2013.
- The underestimation observed in MS was even greater. Compared to the number of prevalent patients identified in 2008-2013, only 79% were identified as being prevalent when the observation period was restricted to 2013 (**Figure 4**).

Figure 4. Underestimation of prevalence estimates depending on different observation periods for DM and MS in 2013 in Germany (reference=observation period 2008-2013)



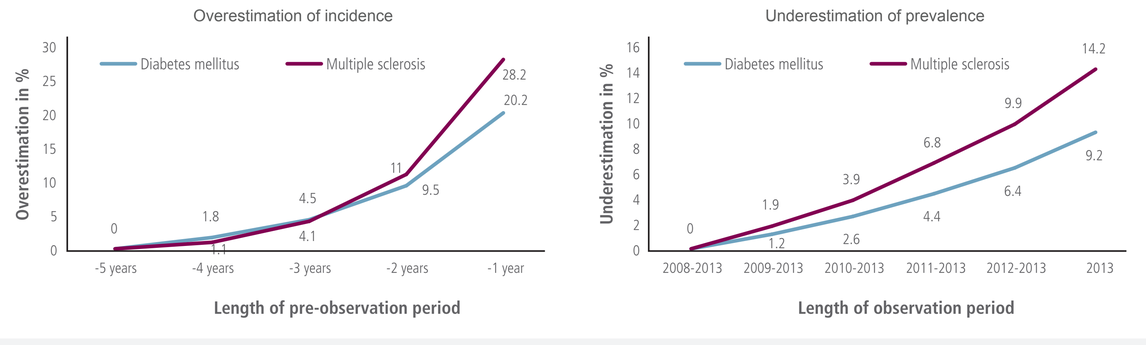
Incidence in Proportion to Prevalence & Change in Proportion

- The proportion of incident cases among prevalent patients varied slightly depending on the estimates used.
- 11% of the prevalent DM patients were identified as incident when the least information available in the HRI Research Database was used, meaning that the observation period was restricted to 2013 for prevalence estimation and that a 1-year pre-observation period was applied for incidence estimation (13% in MS).
- Expanding the pre-observation period for the incidence estimation to 5 years resulted in a relative proportion of 9% incident DM patients among prevalent cases (10% in MS).
- Using all data years available in the HRI Research Database, meaning that an observation period from 2008 to 2013 was applied for prevalence estimation and that a 5-year pre-observation period was applied for incidence estimation, led to a relative proportion of 7% (8% in MS).

Sensitivity Analysis

- The incidence and prevalence were slightly attenuated in both DM and MS when using the stricter case definition.
- Correspondingly, the extent of incidence overestimation and prevalence underestimation was slightly altered. The trend of overestimation and underestimation, however, remained similar (**Figure 5**).

Figure 5. Sensitivity analysis – overestimation of incidence and underestimation of prevalence for DM and MS in 2013 in Germany



CONCLUSIONS

- As shown in previous studies on this issue,[1,2] the pre-observation period should be extended as much as possible to avoid misclassifications when estimating the incidence of chronic conditions based on claims data.
- Nevertheless, the trade-off between prevention of misclassifications and maintaining the representativeness of the sample (eg, in terms of excluding lost to follow-up due to death or change of health insurance) needs to be taken into account when deciding on the length of pre-observation period.
- In addition, the results of this methodological analysis indicate that the observation period in prevalence estimations should also be expanded in order to avoid underestimations.
- The findings were robust over two chronic conditions and during sensitivity analysis.
- Therefore, the methodological concepts should coincide when estimating both the incidence and the prevalence of chronic diseases in claims data.

1. Abbas S, Ihle P, Köster I, Schubert I. Estimation of disease incidence in claims data dependent on the length of follow-Up: a methodological approach. *Health Services Research*. 2012;47(2):746-55.
2. Griffiths RI, O'Malley CD, Herbert RJ, Danese MD. Misclassification of incident conditions using claims data: impact of varying the period used to exclude pre-existing disease. *BMC Medical Research Methodology*. 2013;13(1):32.
3. DM: ATC code A10 (antidiabetics)
MS: ATC codes L03AB07 (Interferon -1a), L03AB08 (Interferon -1b), L03AX13 (Glatiramer acetate), L04AA23 (Natalizumab), L04AA27 (Fingolimod), N07XX09 (Dimethyl fumarate), L04AA13 (Teriflunomide), L01DB07 (Mitoxantrone), N07XX07 (Fampridine), L01AA01 (Cyclophosphamide), L04AX01 (Azathioprin), L01XC04 (Alemtuzumab), H02AB04 (Methylprednisolone), H02AB02 (Dexamethasone), H02AB06 (Prednisolone), H02AB07 (Prednisone), H01AA (ACTH)

Presented at ISPOR 18th Annual European Congress | Milan, Italy | 7-11 November 2015

PRM226