TEMPORAL ORDER OF DISEASE PAIRS AFFECTS SUBSEQUENT DISEASE TRAJECTORIES: THE CASE OF DIABETES AND SLEEP APNEA

METTE K. BECK
Novo Nordisk Foundation Center for Protein Research,
University of Copenhagen, Blegdamsvej 3B
Copenhagen, DK-2200, Denmark
Email: mette.beck@cpr.ku.dk

DAVID WESTERGAARD
Novo Nordisk Foundation Center for Protein Research,
University of Copenhagen, Blegdamsvej 3B
Copenhagen, DK-2200, Denmark
Email: david.westergaard@cpr.ku.dk

ANDERS BOECK JENSEN
Novo Nordisk Foundation Center for Protein Research,
University of Copenhagen, Blegdamsvej 3B
Copenhagen, DK-2200, Denmark
Email: anders.b.jensen@cpr.ku.dk

LEIF GROOP
Lund University Diabetes Centre, Department of Clinical Sciences,
Jan Waldenströms gata 35, SE-205 02 Malmö, Sweden
Email: Leif.Groop@med.lu.se

SØREN BRUNAK
Novo Nordisk Foundation Center for Protein Research,
University of Copenhagen, Blegdamsvej 3B
Copenhagen, DK-2200, Denmark
Email: soren.brunak@cpr.ku.dk

Most studies of disease etiologies focus on one disease only and not the full spectrum of multimorbidities that many patients have. Some disease pairs have shared causal origins, others represent common follow-on diseases, while yet other co-occurring diseases may manifest themselves in random order of appearance. We discuss these different types of disease co-occurrences, and use the two diseases "sleep apnea" and "diabetes" to showcase the approach which otherwise can be applied to any disease pair. We benefit from seven million electronic medical records covering the entire population of Denmark for more than 20 years. Sleep apnea is the most common sleep-related breathing disorder and it has previously been shown to be bidirectionally linked to diabetes, meaning that each disease increases the risk of acquiring the other. We confirm that there is no significant temporal relationship, as approximately half of patients with both diseases are diagnosed with diabetes first. However, we also show that patients diagnosed with diabetes before sleep apnea have a higher disease burden compared to patients diagnosed with sleep apnea before diabetes. The study clearly demonstrates that it is not only the diagnoses in the patient's disease history that are important, but also the specific order in which these diagnosis are given that matters in terms of outcome. We suggest that this should be considered for patient stratification.

1. Introduction

Much epidemiological research has focused on simple associations between two diseases. Temporal approaches have been suggested to uncover both causal and genetic links among statistically associated diseases [1-4]. Many recent studies have analyzed more complicated relations between several diseases and have found bidirectional relationships, where one disease increases the risk or severity of the other or vice versa [1–4]. This type of relationship is mostly found for pairs of common diseases such as depression, cardiovascular diseases and diabetes [2,4]. In one example Mezuk et al. reported a 15% increased risk of depression in patients with type 2 diabetes (T2D), but 60% increased risk of developing type 2 diabetes in patients with depression [5]. Since then several papers have confirmed this particular bidirectional observation [6,7]. Similarly, diabetes has been bidirectionally linked with both periodontitis and sleep apnea [1,8,9].

Until now there has not been general studies investigating the effect of the temporal order in which bidirectionally linked diseases are diagnosed, and how the order affects the further disease progression and the general health status of the patients. In this study we highlight the importance of the temporal order using the bidirectionally linked disease pair: diabetes and sleep apnea. Subsequently we generalize this method to a disease-spectrum wide approach for T2D patients.

Sleep apnea is the most common sleep-related breathing disorder, affecting up to 10% of middle-aged women and up to 20% of middle-aged men in high-income and Asian countries [10–12]. It is traditionally stratified into obstructive sleep apnea and central sleep apnea, where obstructive sleep apnea is the most prevalent subgroup that accounts for up to 85% of sleep apnea patients [13–15]. Furthermore, sleep apnea can occur in both children and adults, although these are treated as two different diseases [16–19]. Untreated, sleep apnea increases the risk for cardiovascular, metabolic, and neurocognitive complications and it is therefore a prototypical example of a disease involved in comorbidities [20,21]. Specifically, it is associated with T2D [1,9,22,23].

Although obesity is a predictor of both obstructive sleep apnea and T2D, the bidirectional link between these diseases appears to be independent of weight [1,9,20]. T2D contributes to sleep apnea, by causing neuromyopathy, which impairs reflexes of the upper airway [9,20]. Sleep apnea contributes to the development of T2D by increased activation of the sympathetic nervous system leading to increased insulin resistance [22,24,25]. It has even been suggested that successful treatment of sleep apnea may reduce the risk of T2D, although this is still controversial [9].

To investigate the effect of the order of the diagnoses we combined the Danish National Patient Registry (NPR), which covers all hospital encounters, both public and private, in Denmark from 1994 to 2015, a patient population of nearly seven million individuals with prescription data from the Danish diabetes registry. NPR records diseases using the International Classifications of Diseases, 10th revision (ICD-10), which organizes diseases hierarchically.

Using this unbiased, country-wide data set we describe the comorbidity map of sleep apnea patients in a data driven manner, and show that the diagnostic order of sleep apnea and T2D is close to 50:50. Interestingly, while the order overall appears to be random we show that the order is associated with significantly different frequencies of comorbid diseases, implying two distinct patient groups.

T2D is a chronic disease with a high risk of many servere complications, including cardiovascular, neurological and infectious complications [5,6,26–30]. Consequently, we generalized our approach to all diseases appearing together with T2D. We showed that the disease burden was dependent on the diagnosis order for twelve T2D comorbidities, of which ten show an increase in comorbidities if T2D was diagnosed first.

2. Materials and methods

In this retrospective cohort study we investigated the association between sleep apnea and T2D. We used the NPR, covering all hospital encounters in Denmark from 1994 to 2015, from where we could include 6,923,849 Danish subjects. Specifically, this registry contained 218,750 T2D patients and 95,853 sleep apnea patients.

To define T2D patients we combined the NPR with the Danish Diabetes registry, which contains medical prescription data. We defined T2D patients, as patients diagnosed at least two times with NIDDM but not with IDDM, if oral hypoglycemic agents were prescribed at least two times and they were diagnosed with NIDDM, or if oral hypoglycemic agents and insulin were prescribed at least two times and they were diagnosed with NIDDM and/or IDDM.

Adult sleep apnea patients were defined as patients first diagnosed with sleep apnea at the age of 16 years or older.

2.1. *Comorbidity calculations*

We tested for significant associations between all level three diagnoses in the ICD-10 terminology. The relative risk of a particular disease was calculated using the Cochran–Mantel–Haenszel method, where each bin corresponds to patients of a particular gender and born in a particular decade. We included patients born from 1900 until 2015, giving rise to up to 24 bins per test. We used the Cochran–Mantel–Haenszel test to identify the p-value and accepted results with

a Benjamini-Hochberg corrected p-value of 0.05 or below. This method was used both for time-independent and time-dependent analyses.

2.2. *From temporal diagnosis pairs through disease trajectories to disease network*

The method for identifying the trajectories has been described previously in detail [33]. The method consists of three steps: First temporal directed pairs of co-morbid diseases were tested to identify pairs where one disease is associated with an increase in the occurrences of the other. In the second step, the pairs found are tested for directionality (one disease primarily occurring before the other) using a binomial test. Third, the pairs with significant temporality were combined into disease trajectories of three consecutive diseases. Trajectories were only included if at least 100 sleep apnea patients followed them.

2.3. *Difference in mean number of comorbidities*

The difference in mean number of comorbidities was modeled by a Poisson regression using the covariates: years between the two diagnoses, which disease was diagnosed first, age and gender. All four covariates significantly contributed to the model. This Poisson regression was subsequently used to predict the number of comorbidities for all patients to avoid age and/or gender bias. The difference in mean predicted number of comorbidities was tested using student's t-test, stratified by the order of the diagnoses. This was done twenty times, requiring a minimum from zero years up to nineteen years in between the two diagnoses.

2.4. *Diabetes comorbidities selection criteria*

We tested if any level three ICD-10 diagnoses were significantly correlated with T2D using the method for comorbidity calculations. For the diagnosis with a significant association and a relative risk above one, we used a binomial test to ensure lack of directionality. We required the 95% confidence interval to be within 45% - 55% (making the diagnostic order close to 50:50). Lastly, we required a minimum of 1,000 T2D patients to have the disease. For the remaining diseases we performed the method described in "Difference in mean number of comorbidities". We required ten time points to be significant.

3. Results

In the Danish population of 6,923,707 patients, we found 117,913 patients diagnosed with sleep disorders (G47), of these 95,853 patients were diagnosed with sleep apnea (G47.3). The age distribution at which these patients were first diagnosed with sleep apnea is shown in Figure 1A. It has two clearly distinct peaks, the first at age three, and the second major peak just after 50 years, supporting that this diagnosis could cover two distinct disease progression patterns. We computed the relative risk (RR) for both the adult onset of sleep apnea (aged 16 or above) and the childhood onset, compared to all other level three ICD-10 diagnoses. Even though both groups of patients are diagnosed with the same diagnosis, their repertoire of comorbidities is very different (Figure 1B), in part due to the difference in age. We therefore excluded childhood onset of sleep apnea, and

investigated sleep apnea in the adult population further. Of the 95,853 sleep apnea patients 90,157 were diagnosed in adult patients, 75% of these were males.

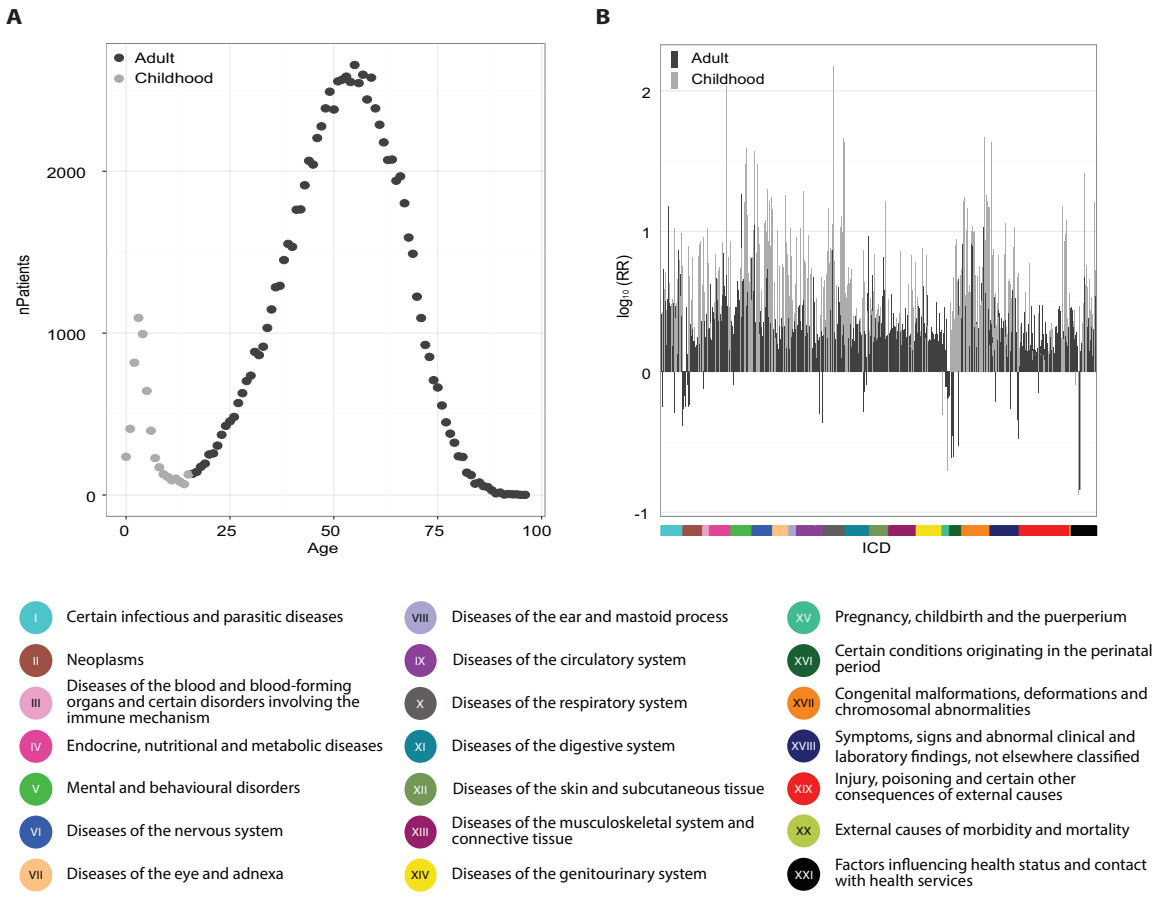

Fig. 3. The increased comorbidity burden for patients diagnosed with T2D before sleep apnea. (A) Distribution of years between T2D and sleep apnea for patients diagnosed with T2D first (pink) and for patients diagnosed with sleep apnea first (blue). (B) The excess number of comorbidities for patients diagnosed with T2D first compared to those diagnosed with sleep apnea first (black line) with the 95% confidence interval (grey area). The x-axis indicates the minimum number of whole years between diagnoses (e.g. 0 years means more than one day but less than a year). The dots indicate the number of patients having minimum x years between the two diagnoses.

3.1. *Temporal disease network reveals no direct connection between diabetes and adult sleep apnea*

We identified all diseases that co-occurred more often than we would expect from their individual frequencies in the patients with adult sleep apnea. For each such disease pair, we testedif one of the diseases occurred significantly more often before the other. This led to the identification of a pool of significant, directed disease-pairs (see Methods). These pairs were combined into linear, temporal disease trajectories of which we found 103 where 100 sleep apnea

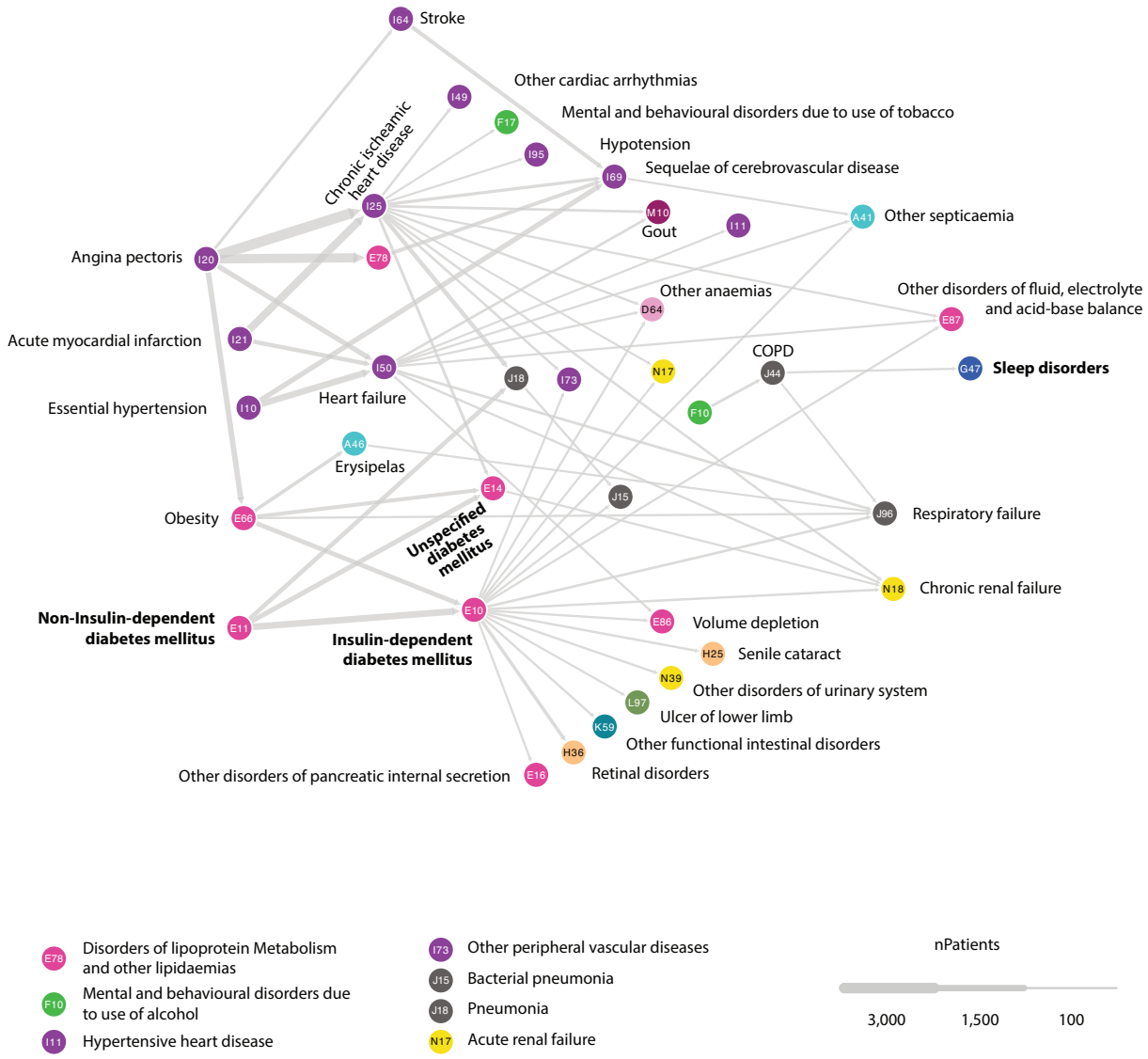

Fig. 2. Temporal disease network based on sleep apnea patients. The network was constructed from 103 sleep apnea trajectories and illustrates the number of patients taking a particular step in the disease network (width of arrow). The nodes are colored based on their ICD-10 chapter relationships. The names are written next to their node in the network or mentioned in the legend in alphabetical order.

patients followed three consecutive steps of diseases. Subsequently, the 103 linear trajectories found in the adult sleep apnea patient group were combined into a temporal disease network providing a concerted overview of the comorbidity spectrum (Figure 2). As expected, obesity, a known risk factor for sleep apnea, appears as a statistically significant component in this overview network (present in 20 of the 103 temporal trajectories as either starting or midpoint). Several cardio-vascular complications are also prominent in the network. Additionally, both insulin-dependent diabetes mellitus (IDDM) and non-insulin-dependent diabetes mellitus (NIDDM) are part of the disease network along with several diabetes complications. There is no direct path

connecting diabetes and sleep disorders in the disease network, due to the lack of temporality between these diagnoses.

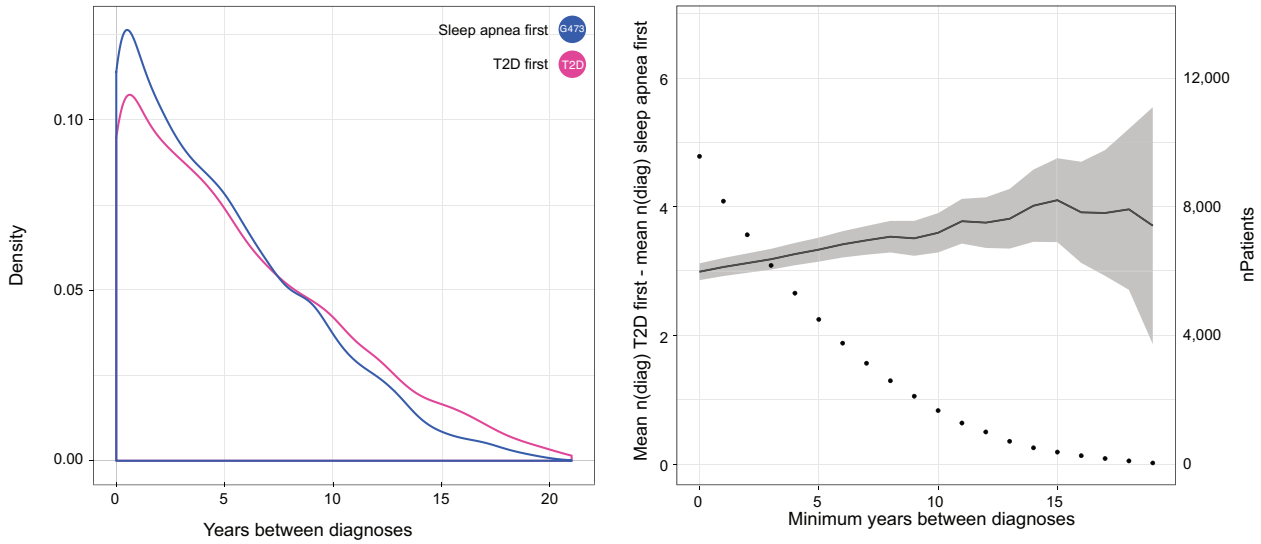

Fig. 3. The increased comorbidity burden for patients diagnosed with T2D before sleep apnea. (A) Distribution of years between T2D and sleep apnea for patients diagnosed with T2D first (pink) and for patients diagnosed with sleep apnea first (blue). (B) The excess number of comorbidities for patients diagnosed with T2D first compared to those diagnosed with sleep apnea first (black line) with the 95% confidence interval (grey area). The x-axis indicates the minimum number of whole years between diagnoses (e.g. 0 years means more than one day but less than a year). The dots indicate the number of patients having minimum x years between the two diagnoses.

3.2. *Diabetes before sleep apnea is associated with an increased amount of comorbidities*

To further investigate the temporal association between sleep apnea and T2D, we defined T2D patients based on the method presented by Lind at al [29,31], using a combination of prescribed drugs and disease codes (see Methods). We found that 11,054 T2D patients have been diagnosed with sleep apnea. A total of 6,061 patients (54,8%) were diagnosed with T2D before sleep apnea, and 4,752 patients were diagnosed with sleep apnea before T2D. In addition, 241 patients were diagnosed with T2D and sleep apnea on the same day. These 241 patients are disregarded in this study, since there is no reliable way to determine which disease came first. Consequently, even though sleep apnea was significantly associated with T2D (RR = 2.87, $p < 2.3E\text{-}308$), NIDDM and sleep apnea does not appear as a temporal pair, due to the lack of a significant temporal order in which these diseases are diagnosed.

To investigate if the patients diagnosed with T2D before adult sleep apnea and patients diagnosed with adult sleep apnea before T2D are two distinct patient groups, we examined the RR for all level three ICD-10 diagnoses for patients with adult sleep apnea and T2D. Those first diagnosed with T2D had on average 3.0 (95% CI: 2.9-3.1) comorbidities more than those

diagnosed with adult sleep apnea first. We interpret this as an indicator that the patients first diagnosed with T2D have, on average, a higher disease burden.

The time of diagnosis can be imprecise since neither sleep apnea nor T2D are acute diseases. Consequently, it could be arbitrary which disease was diagnosed first. For some patients the two diagnoses are acquired relatively close to each other, but for many patients there are several years or even decades between the diagnoses (Figure 3A).

We tested if there was a significant difference in the number of diagnoses between these two groups using a Poisson regression model. Covariates include years between the two diagnoses, which disease was diagnosed first, age and gender. We used the fitted model to calculate a point estimate of the number of comorbidities for each patient, given the minimum number of years between sleep apnea and T2D (Figure 3B). The overall difference was 3.0 comorbidities, with patients first diagnosed with T2D being most sick. This difference increases as the number of years between T2D and sleep apnea increases (Figure 3B). Collectively, this clearly illustrates a difference in the general health status of these patients groups.

3.3. *Diabetes before other diseases tends to increase the comorbidity burden*

We applied the same method to investigate if other diabetes comorbidities showed a different comorbidity burden depending on the diagnosis order. We found seventeen diseases positively associated with T2D, and where the diagnostic order for each disease and T2D was close to 50:50 (see Methods). To remove rare disorders we required a minimum of 1,000 T2D patients to have the diagnosis, reducing the number down to sixteen diagnoses of interest. Lastly, we performed an analysis calculating the difference in mean number of comorbidities for patients diagnosed with T2D first compared to patients diagnosed with the other particular diagnosis first. This resulted in twelve diagnoses with a minimum of ten significant time points (Figure 4). Ten out of the twelve diagnoses were associated with a higher comorbidity burden if they were diagnosed with T2D before the other diagnosis, with the two exceptions: Migraine and “Poisoning by psychotropic drugs, not elsewhere classified”.

4. Discussion

In this study we examined the complex issue of temporal directionality of disease co-occurrences and used temporal disease trajectories to present a model for stratification of patient groups according to longitudinal patterns.

Using one example analyzed in detail we illustrated the complexities and rediscovered that age of sleep apnea diagnoses follow a bimodal distribution, illustrating two distinct diseases: childhood sleep apnea and adult sleep apnea – a distinction well known in the literature [11,16–18,32]. By investigating the detailed time-ordered relationships between sleep apnea and T2D we confirmed that sleep apnea in the adult population is significantly associated with T2D in the time-dependent analysis. Surprisingly, there was no direct edge between any of the diabetes diagnoses in our temporal disease network, showing that there was no directionality of the T2D and adult sleep apnea diagnoses, in fact we showed that 4,752 patients acquire adult sleep apnea before T2D, 6,061 acquire T2D first while 241 patients acquired the diagnoses on the same day.

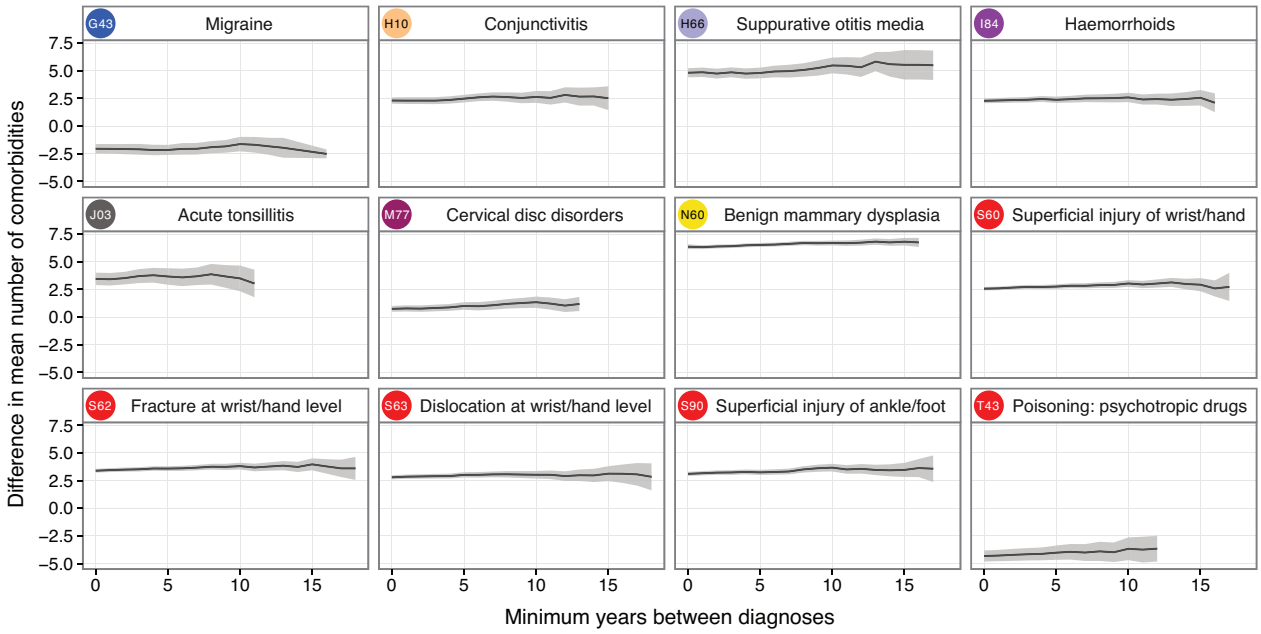

Fig. 4. Comorbidity burden levels as function of time span between diagnoses. Together the panels show that the change in comorbidity burden depends on the disease order. Each disease is indicated by an ICD-10 code colored according to the ICD-10 chapter followed by the name of the disease. The excess number of comorbidities for patients diagnosed with T2D first compared to those diagnosed with the other particular diagnosis (black line) with the 95% confidence interval (grey area). The x-axis indicates the minimum number of whole years between diagnoses.

Interestingly, we found that this order significantly influenced the amount of comorbidities acquired, indicating that patients diagnosed with diabetes before adult sleep apnea have a worse general health status than patients first diagnosed with adult sleep apnea. This is, to our knowledge, the first time this temporal effect of sleep apnea and T2D has been described. To further illustrate the importance of the order, we showed that the difference in the quantity of comorbidities slightly increased with increased time between the diagnoses. Based on these observations we suggest that there is a synergetic effect of T2D and adult sleep apnea, which is dependent on the order of the diagnoses.

We further underlined the importance of order of diagnoses by applying this method to all T2D comorbidities. This resulted in twelve diagnoses with a significant different number of comorbidities depending on the diagnosis order.

Precision medicine attempts to subdivide patients into groups that will benefit from tailor-made treatment. We show in this paper that disease progression patterns can be highly complex even in cases where disease co-occurrence orders appear to be random. The identification of genomic biomarkers could most likely to a higher degree benefit from taking this type of stratification into account in contrast to current models that mostly are based on the case/control paradigm where diseases are investigated individually.

5. Acknowledgements

We would like to acknowledge funding from the Novo Nordisk Foundation (grant agreement NNF14CC0001) as well as the H2020 project MedBioInformatics.

References

1. Aurora, R. N. & Punjabi, N. M. *Lancet Respir. Med.* **1,** 329–338 (2013).
2. Golden, S. H. *et al. JAMA* **299,** 2751–2759 (2008).
3. Hesdorffer, D. C. *et al. Ann. Neurol.* **72,** 184–191 (2012).
4. Lippi, G., Montagnana, M., Favaloro, E. J. & Franchini, M. *Seminars in Thrombosis and Hemostasis* **35,** 325–336 (2009).
5. Mezuk, B., Eaton, W. W., Albrecht, S. & Golden, S. H. *Diabetes Care* **31,** 2383–2390 (2008).
6. Pan, A. *et al. Arch. Intern. Med.* **170,** 1884–91 (2010).
7. Pan, A. *et al. Diabetes Care* **35,** 1171–1180 (2012).
8. Lalla, E. & Papapanou, P. N. *Nat. Rev. Endocrinol.* **7,** 738–48 (2011).
9. Rajan, P. & Greenberg, H. *Nat. Sci. Sleep* **7,** 113–25 (2015).
10. Peppard, P. E. *et al. Am. J. Epidemiol.* **177,** 1006–1014 (2013).
11. Sharma, S. K. & Ahluwalia, G. *Indian J. of Med. Res.* **131,** 171–175 (2010).
12. Ip, M. S. M. *et al. Chest* **119,** 62–69 (2001).
13. Javaheri, S. *Clinics in Chest Medicine* **31,** 235–248 (2010).
14. Morgenthaler, T. I., Kagramanov, V., Hanak, V. & Decker, P. A. *Sleep* **29,** 1203–1209 (2006).
15. Khan, M. T. & Franco, R. A. *Sleep Disord.* 798487 (2014).
16. Tan, H.-L., Gozal, D. & Kheirandish-Gozal, L. *Nat. Sci. Sleep* **5,** 109–23 (2013).
17. Marcus, C. L. *et al. Pediatrics* **130,** 576–84 (2012).
18. Marcus, C. L. *et al. N. Engl. J. Med.* **368,** 2366–76 (2013).
19. Bixler, E. O. *et al. Sleep* **32,** 731–6 (2009).
20. Malhotra, A. & White, D. P. *The Lancet* **360,** 237–245 (2002).
21. Parati, G. *et al. J. Hypertens.* **30,** 633–46 (2012).
22. Cappuccio, F. P., D'Elia, L., Strazzullo, P. & Miller, M. A. *Diabetes Care* **33,** 414–420 (2010).
23. Malhotra, A. *et al. Am. J. Respir. Crit. Care Med.* **166,** 1388–1395 (2002).
24. Chervin, R. D. *Chest* **118,** 372–379 (2000).
25. Barceló, A. *et al. Thorax* **63,** 946–50 (2008).
26. Fowler, M. J. *Clin. Diabetes* **29,** 116–122 (2011).
27. Alves, C., Casqueiro, J. & Casqueiro, J. *Indian J. Endocrinol. Metab.* **16,** 27 (2012).
28. DeFronzo, R. A. *et al. Nat. Rev. Dis. Prim.* **1,** 15019 (2015).
29. Lind, M. *et al. Diabetologia* **55,** 2946–2953 (2012).
30. Bertoni, A. G., Saydah, S. & Brancati, F. L. *Diabetes Care* **24,** 1044–1049 (2001).
31. Lind, M. *et al. N. Engl. J. Med.* **371,** 1972–1982 (2014).
32. Jordan, A. S., McSharry, D. G. & Malhotra, A. *Lancet* **383,** 736–47 (2014).
33. Jensen, A. B. *et al. Nat. Commun.* **5,** 4022 (2014).