Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study

Ronald B. Postuma,1,2 Alex Iranzo,3 Michele Hu,4 Birgit Högl,5 Bradley F. Boeve,6 Raffaele Manni,7 Wolfgang H. Oertel,8 Isabelle Arnulf,9 Luigi Ferini-Strambi,10 Monica Puligheddu,11 Elena Antelmi,12,13 Valerie Cochen De Cock,14 Dario Arnaldi,15 Brit Mollenhauer,16 Aleksandar Videnovic,17 Karel Sonka,18 Ki-Young Jung,19 Dieter Kunz,20 Yves Dauvilliers,21 Federica Proveni,22,23 Simon J. Lewis,24 Jitka Buskova,25 Milena Pavlova,26 Anna Heidbreder,27 Jacques Y. Montplaisir,2 Joan Santamaría,14 Thomas R. Barber,4 Ambra Stefani,5 Erik K. St. Louis,6 Michele Terzagli,7 Annette Janzen,8 Smanda Leu-Semenescu,9 Guiseppe Plazzi,12,13 Flavio Nobili,15 Friederike Sixel-Doering,16 Petr Dusek,18 Frederik Bes,20 Pietro Cortelli,22,23 Kaylena Ehgoetz Martens,24 Jean-Francois Gagnon,28 Carles Gaig,3 Marco Zucconi,10 Claudia Trenkwalder,15 Ziv Gan-Or,29,30 Christine Lo,4 Michal Rolinski,4 Philip Mahlknecht,5 Evi Holzknecht,5 Angel R. Boeve,6 Luke N. Teigen,6 Gianpaolo Toscano,7 Geert Mayer,31 Silvia Morbelli,32 Benjamin Dawson,1 Amelie Pelletier1,2 and the International REM Sleep Behavior Disorder Study Group

See Morris and Weil (doi:10.1093/brain/awz014) for a scientific commentary on this article.
variables were different at baseline between those converting to primary dementia versus parkinsonism. Sample size estimates for definitive neuroprotective trials ranged from 142 to 366 patients per arm. This large multicentre study documents the high phenoconversion rate from iRBD to an overt neurodegenerative syndrome. Our findings provide estimates of the relative predictive value of prodromal markers, which can be used to stratify patients for neuroprotective trials.

Introduction

The neurodegenerative synuclein aggregation disorders, namely Parkinson’s disease, dementia with Lewy bodies, and multiple system atrophy (MSA), all have a prodromal interval; that is, a period during which neurodegenerative symptoms/signs are present, but full clinical disease has not yet developed (Berg et al., 2015). In the synucleinopathies,
this interval is notably long, often exceeding a decade (Berg et al., 2015). This provides an unprecedented opportunity to provide potential neuroprotective therapy early, perhaps even preventing the development of parkinsonism and dementia.

Unlike many neurological diseases, whose prodromal states are predominantly identified by abnormalities in the same domain [e.g. mild cognitive impairment (MCI) is the primary prodromal marker of Alzheimer’s disease], prodromal synucleinopathy markers are notably diverse. In addition to subtle motor signs, the potential prodromal markers include autonomic abnormalities, olfactory loss, cognitive changes, depression, anxiety, etc. (Goldman and Postuma, 2014). Most are relatively non-specific, such that the large majority of marker-positive subjects will never develop disease. However, a notable exception is idiopathic REM sleep behaviour disorder (iRBD).

iRBD is a parasomnia in which the normal paralysis of REM sleep is lost, such that patients ‘act out’ their dreams (Schenck et al., 2013b; Hogl et al., 2018). Idiopathic RBD [alternatively termed ‘isolated’ (Hogl et al., 2018) or ‘cryptogenic’ RBD] has a prevalence of ~1% over age 60 (Kang et al., 2013; Haba-Rubio et al., 2018; Pujol et al., 2017), although most do not present to medical attention. Observational studies, generally from single centres, have suggested that most patients with iRBD will eventually develop a defined neurodegenerative disease, almost always diagnosed as synucleinopathy (Wing et al., 2012; Schenck et al., 2013a; Irazo et al., 2014; Arnulf et al., 2015; Mahlknecht et al., 2015; Postuma et al., 2015a, d; Li et al., 2017). In this context, RBD is likely related to neurodegeneration in the pontine or medullary areas associated with control of REM atonia (Valencia Garcia et al., 2018). The latency from symptom onset to disease phenoconversion (i.e. conversion from iRBD to defined dementia with Lewy bodies, Parkinson’s disease, or MSA) averages over 10 years (Schenck et al., 2013b). Therefore, this implies that 1% of the elderly population have a readily-diagnosable but often-undetected early-stage neurodegenerative syndrome.

So far, most studies of phenoconversion risk and predictors came from single centres, so whether this is seen across different countries and different contexts remains unclear. In this study, we combined the prospective experience of 24 centres from the International RBD Study Group, to quantify the risk of phenoconversion to defined parkinsonism/dementia and to test 21 potential predictors of phenoconversion.

**Materials and methods**

**Subjects**

For inclusion, all subjects had to have iRBD confirmed on polysomnogram according to American Academy of Sleep Medicine Criteria (American Academy of Sleep Medicine and Hauri, 2007), and be free of parkinsonism or dementia on baseline neurological examination. Each patient had at least one follow-up examination during which systematic assessment for parkinsonism and dementia was performed. All patients gave written informed consent according to the Declaration of Helsinki, and ethics approval was obtained from the local institutional boards.

**Baseline variables**

Centres collected all available information on baseline variables, then followed patients prospectively. We did not require that each variable be tested in each patient; rather, centres sent results for all those variables that they systematically assessed. Neither did we require that all variables be assessed with the same technique, as centres had different testing protocols for prodromal markers. For the analyses of hazard ratio (HR) in with tests were categorized as abnormal or normal, each centre defined each variable as abnormal/normal according to their own testing protocols, unless otherwise stated below. Detailed numbers of patients assessed with each variable is provided in Supplementary Table 1. Variables of interest and the assessment methods used included:

(i) Standardized motor examination: tested with the Movement Disorders Society Unified Parkinson Disease Rating Scale (MDS-UPDRS) (Goetz et al., 2008). Either the 1987 UPDRS or 2008 MDS-UPDRS version could be used. For stratification purposes, the cut-off score was >3 excluding action tremor (Postuma et al., 2012).

(ii) Standardized motor symptoms: UPDRS/MDS-UPDRS Part II (Fahn et al., 1987; Goetz et al., 2008).

(iii) Quantitative motor testing: tests included the alternate-tap test (Schnurr et al., 2000; Postuma et al., 2015); Purdue Pegboard (Desrosiers et al., 1995; Postuma et al., 2015); 3-Metre Timed-Up-and-Go (Podsiadlo and Richardson, 1991; Postuma et al., 2015), or Flamingo balance test (Barber et al., 2017). If multiple tests were conducted in one centre, the majority had to be abnormal to classify the testing as abnormal.

(iv) Olfaction: 12- or 40-item University of Pennsylvania Smell Identification Test (Doty et al., 1984) or Sniffin Sticks (Hummel et al., 1997; Mahlknecht et al., 2015).

(v) Colour vision: Farnsworth-Munsell 100-Hue test (Farnsworth, 1943).

(vi) Physician-documented insomnia: Insomnia Severity Index (Bastien et al., 2001), Athens Insomnia Scale (Soldatos et al., 2000), or clinical interview.

(vii) Excessive daytime somnolence: Epworth Sleepiness scale (Johns, 1991; Valencia Garcia et al., 2018) or clinical interview.

(viii) Restless legs syndrome: diagnosed using clinical interview.

(ix) Sleep apnoea: apnoea-hypopnoea index cut-off ≥ 15/h (second- ary analysis was also performed using cut-off ≥ 5/h).

(x) REM sleep without atonia: scored as % tonic and phasic chin REM on the polysomnographic trace, using either Montreal scoring (Montplaisir et al., 2010), or % ‘any’ tone using SINBAR scoring, chin + arm (Frauscher et al., 2012). For combined stratification, we divided each individual’s score by the mean estimate from their centre.

(xi) Constipation: Unified MSA Rating Scale (UMSARS) (Wenning et al., 2004), SCOPA-AUT (Visser et al., 2004), Rome Criteria (Higgins and Johanson, 2004), or clinical interview.

(xii) Urinary symptoms: UMSARS, SCOPA-AUT, or clinical interview.
Follow-up and disease conversion

All centres prospectively followed patients with in-person evaluation to diagnose phenoconversion to defined parkinsonism [defined as bradykinesia plus at least one of rigidity or rest tremor (Postuma et al., 2015b)] or dementia [defined as cognitive impairment on standardized testing with functional impairments (Postuma et al., 2006)] or dementia [defined as cognitive impairment on standardized testing with functional impairments (Postuma et al., 2006)] or dementia [defined as cognitive impairment on standardized testing with functional impairments (Postuma et al., 2006)] or dementia [defined as cognitive impairment on standardized testing with functional impairments (Postuma et al., 2006)]. For patients with parkinsonism as the primary disease manifestation, the primary diagnosis (Parkinson’s disease/MSA) was made according to the treating neurologist. This differential diagnosis incorporated all available follow-up information (i.e. any patient who was initially diagnosed with Parkinson’s disease at phenoconversion but who was subsequently found to have MSA would be included as MSA). For dementia conversions, all patients had polymorphonuclear white blood cell counts greater than 10,000/mL and no other concomitant illnesses. For those patients who developed dementia as the first disease manifestation or both variables present) were eligible for combined analysis.

Finally, we estimated sample size requirements for a future neuroprotective trial. This assumed a categorical definitive end-point (defined disease phenoconversion), with groups (placebo versus a single-dose of active treatment), two-sided alpha = 0.05, and 80% power. We used time-to-event analysis (http://www.quesgen.com/SSSurvival.php), for a 2-year trial, assuming an agent that reduces phenoconversion with HR = 0.5. We calculated sample size for the population as a whole, using stratification by prodromal marker testing, using directly-observed conversion rates, and also by using the hazard ratio from the current study estimates (i.e. adjusting for centre effects by recalculating the conversion rate in each single analysis to equal the median conversion rate in the entire group). For assessment of MDS prodromal criteria, we included only patients who had sufficient testing to reasonably estimate their 3% probability, which was defined as four or more prodromal variables including at least one of the three highest-specificity variables (olfaction, objective motor examination/quantitative testing, DAT-SPECT); for all calculations, the likelihood ratio of RBD (130) was included.

Data availability

The original database from the study can be obtained by contacting the first author (R.B.P.).

Results

Participants

A total of 1280 patients from 24 centres were included in this study. Recruitment data from each centre are
<table>
<thead>
<tr>
<th>Centre</th>
<th>n</th>
<th>Follow-up duration, y, mean</th>
<th>Total patient-years</th>
<th>UPDRS MDS-UPDRS Part III</th>
<th>Quantitative motor testing</th>
<th>UPDRS MDS-UPDRS Part II</th>
<th>Olfaction vision</th>
<th>Colour Other sleep symptoms</th>
<th>PSG - apnoea</th>
<th>PSG - REM atonia %</th>
<th>Autonomic symptoms</th>
<th>Orthostatic blood pressure</th>
<th>Office-based cognitive tests</th>
<th>Neuropsychological examination</th>
<th>Depression</th>
<th>Anxiety</th>
<th>Substantia nigra ultrasound</th>
<th>DAT-SPECT T total patient/year/variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Montreal</td>
<td>154</td>
<td>4.3</td>
<td>664</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>10789</td>
</tr>
<tr>
<td>Barcelona</td>
<td>202</td>
<td>5.4</td>
<td>1091</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>3926</td>
</tr>
<tr>
<td>Oxford</td>
<td>120</td>
<td>2.2</td>
<td>266</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>3790</td>
</tr>
<tr>
<td>Innsbruck</td>
<td>98</td>
<td>4.5</td>
<td>461</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>3559</td>
</tr>
<tr>
<td>Rochester, Minnesota</td>
<td>96</td>
<td>4.4</td>
<td>424</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>3434</td>
</tr>
<tr>
<td>Pavia</td>
<td>74</td>
<td>5.4</td>
<td>400</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>3240</td>
</tr>
<tr>
<td>Marburg</td>
<td>43</td>
<td>3.4</td>
<td>145</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>1697</td>
</tr>
<tr>
<td>Paris</td>
<td>97</td>
<td>2.9</td>
<td>278</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>1585</td>
</tr>
<tr>
<td>Milan</td>
<td>51</td>
<td>2.4</td>
<td>123</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>1583</td>
</tr>
<tr>
<td>Cagliari</td>
<td>37</td>
<td>2.9</td>
<td>108</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>1295</td>
</tr>
<tr>
<td>Bologna (Antelme)</td>
<td>24</td>
<td>4.2</td>
<td>100</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>990</td>
</tr>
<tr>
<td>Montpellier (Cadouin)</td>
<td>34</td>
<td>1.9</td>
<td>65</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>916</td>
</tr>
<tr>
<td>Genoa</td>
<td>31</td>
<td>2.7</td>
<td>84</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>856</td>
</tr>
<tr>
<td>Kassel</td>
<td>22</td>
<td>2.3</td>
<td>50</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>795</td>
</tr>
<tr>
<td>Boston (MGH)</td>
<td>16</td>
<td>3.4</td>
<td>54</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>680</td>
</tr>
<tr>
<td>Prague (Sonka)</td>
<td>31</td>
<td>1.5</td>
<td>46</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>662</td>
</tr>
<tr>
<td>Seoul</td>
<td>28</td>
<td>2.1</td>
<td>58</td>
<td>±</td>
<td>±</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>557</td>
</tr>
<tr>
<td>Berlin</td>
<td>47</td>
<td>1.2</td>
<td>55</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>528</td>
</tr>
<tr>
<td>Montpellier (Dauvillers)</td>
<td>17</td>
<td>3.5</td>
<td>60</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>281</td>
</tr>
<tr>
<td>Bologna (Provin)</td>
<td>9</td>
<td>2.3</td>
<td>21</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>277</td>
</tr>
<tr>
<td>Sydney</td>
<td>13</td>
<td>2.2</td>
<td>29</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>200</td>
</tr>
<tr>
<td>Prague</td>
<td>9</td>
<td>1.3</td>
<td>12</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>162</td>
</tr>
<tr>
<td>Boston (Buskova)</td>
<td>6</td>
<td>4.0</td>
<td>24</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>130</td>
</tr>
<tr>
<td>Muenster</td>
<td>19</td>
<td>1.2</td>
<td>23</td>
<td>X</td>
<td>±</td>
<td>X</td>
<td>X</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>124</td>
</tr>
</tbody>
</table>

| Total                  | 1280| 72.7                       | 4890                | ±                         | ±                         | ±                         | ±                | ±                         | ±             | ±                | ±                   | ±                          | ±                           | ±                             | ±                     | ±       | ±                   | 42698                          |
summarized on Table 1. Mean age at baseline was 66.3 ± 8.4 and 82.5% were male. The mean follow-up duration (between first baseline examination and last contact or disease conversion) was 3.6 years (maximum = 19 years), translating to 4890 total person-years of follow-up.

### Overall outcome

During follow-up, 352 (28%) converted to an overt neurodegenerative syndrome (Fig. 1). The mean interval between baseline examination and phenoconversion was 4.6 ± 3.5 years. The median time to phenoconversion was 8.0 years, with a mean age at baseline of 67.6 years. The mean interval between first baseline examination and last contact or disease conversion was 3.6 years (maximum = 19 years), translating to 4890 total person-years of follow-up.
with an overall phenoconversion rate of 6.25% per year. The risk of phenoconversion on Kaplan-Meier analysis was 10.6% after 2 years, 17.9% after 3 years, 31.3% after 5 years, 51.4% after 8 years, 60.2% after 10 years, and 73.5% after 12 years. With regards to disease classifications, 199 (56.5%) developed parkinsonism as the first disease manifestation [of whom 16 (4.5%) were diagnosed with probable MSA], and 153 (43.5%) developed dementia first.

Predictors of outcome
Kaplan-Meier analysis of selected predictors is illustrated on Fig. 2. On Cox proportional hazards analysis, adjusting for age, sex, and centre, numerous measures significantly predicted outcome (Table 2 and Figs 2–4). These included: (i) quantitative motor testing (HR = 3.16); (ii) standardized motor examination [HR = 3.03 overall, higher for MDS-UPDRS (3.77) than UPDRS-III (2.75)]; (iii) olfaction (HR = 2.62). Predictive value also improved when excluding MSA patients (HR = 2.91); (iv) MCI, with better prediction using neuropsychological examination (HR = 2.37) than with office-based testing (HR = 1.91); (v) erectile dysfunction (HR = 2.13); (vi) motor symptoms: HR = 2.11, with better prediction for the MDS-UPDRS-II (HR = 4.75) than the 1987 UPDRS-II (HR = 1.29); (vii) DAT-SPECT (HR = 1.98); (viii) neuropsychological testing (regardless of cognitive complaint) (HR = 1.89); (ix) constipation (HR = 1.69); (x) sleep (HR = 1.67); (xi) REM sleep without atonia (HR = 1.54, on combined analysis only); (xii) brief office-based cognitive tests (regardless of cognitive complaint) (MMSE/ MoCA combined HR = 1.55); and (xiii) age (HR = 1.54 for above versus below mean).

In addition, systolic blood pressure drop at a cut-off of 10 mm (HR = 1.55) predicted outcome on unadjusted analysis, but not after adjusting for age, sex, and centre (HR = 1.37) (using a cut-off of 20 mm, the unadjusted HR was 1.37 (0.88–2.15) and adjusted HR was 1.20 (0.74–1.91). The MDS prodromal criteria (which combines numerous variables) predicted outcome with the highest hazard ratio (HR = 5.37).

By contrast, we saw no significant predictive differences according to sex, insomnia symptoms, daytime somnolence, restless legs syndrome, apnoea, urinary dysfunction, orthostatic symptoms, depression, anxiety, or substantia nigra ultrasound.

Secondary and sensitivity analyses
Among the 336 patients diagnosed with Lewy Body disease (i.e. excluding MSA), there were relatively few differences between patients who converted to dementia first versus parkinsonism first (Table 3). Age and sex were similar. All motor measures were similar except for quantitative motor testing, which was more likely to be abnormal in those developing dementia first (82.4%) than parkinsonism first (47.2%). Olfaction was similar in both groups, as were all sleep symptoms and polysomnographic variables. Autonomic symptoms were similar, as was orthostatic blood pressure drop, depression or anxiety. Although power was limited, we also saw no differences in proportion of patients with abnormal DAT-SPECT or substantia nigra ultrasound. The only variables that differed strongly (all P < 0.001) were those that tested cognition, including office based cognitive testing, neuropsychological examination, and colour vision testing which predicted only dementia [note that colour vision predominantly tests visuoperceptual cognition in Parkinson’s disease (Bertrand et al., 2012)].

Excluding results from centres that already published data on these predictors did not substantially affect the hazard ratio. For example, the hazard ratio of UPDRS excluding Montreal (Postuma et al., 2012) was 3.04, versus 3.03 for entire group. The hazard ratio of olfaction excluding both Montreal (Postuma et al., 2011) and Innsbruck (Mahlknecht et al., 2015) was 2.53, versus 2.62.

Sample size calculations
Based on the time-to-event analysis, we estimated that 366 patients per arm would need to be recruited into a 2-year trial to have 80% power to find a 50% reduction in disease phenoconversion (i.e. 65 phenoconversion events; Table 4). Adjusting the study duration altered sample sizes roughly proportionally to the proportion in duration (e.g. 4-year trial = 192 per group, 1-year trial = 709 per group). Testing different effectiveness assumptions, a drug providing 80% reduction in phenoconversion would require 84 patients per group (12 phenoconversion events) while a 30% reduction would require 959 (190 phenoconversion events).

The most powerful single selection procedure (abnormal quantitative motor testing) reduced sample size to 166–197 patients; however, only 34% of the iRBD population had abnormal testing and so would be included in such a study.
Figure 2 Kaplan-Meier plot of disease-free survival of patients with iRBD stratified according to presence of motor and cognitive markers. Results are presented according to baseline assessment (i.e. patients who develop a de novo marker abnormality over the course of the follow-up remain in the 'marker-free' group). Solid line indicates patients with normal values, dashed line abnormal values. Hazard ratios (HRs) are with Cox proportional hazards, adjusting for age, sex, and centre, with 95% confidence intervals in parentheses.
Figure 3 Kaplan-Meier plot of disease-free survival of patients with iRBD stratified according to presence of sleep and psychiatric markers. Solid line indicates patients with normal values, dashed line abnormal values. Hazard ratios (HRs) are with Cox proportional hazards, adjusting for age, sex, and centre, with 95% confidence intervals in parentheses.
Figure 4 Kaplan-Meier plot of disease-free survival of patients with iRBD stratified according to presence of special sensory and autonomic markers. Solid line indicates patients with normal values, dashed line abnormal values. Hazard ratios (HRs) are with Cox proportional hazards, adjusting for age, sex, and centre, with 95% confidence intervals in parentheses.
On the other hand, other stratification strategies allowed more inclusions; selecting for abnormal olfaction allowed 67% eligibility with sample size of 247–262 per group, and selecting those who met the MDS prodromal criteria allowed 77% eligibility with sample size of 282–301 per group. Among two-factor combinations, the combination of olfaction and UPDRS retained 29% eligibility, and resulted in an estimated 15.7% annual conversion rate, translating into 157 patients per group.

**Discussion**

In this large multicentre study, we have confirmed the very high risk of Parkinson’s disease, dementia with Lewy bodies, and MSA in ‘idiopathic’ RBD, and have confirmed numerous predictors of outcome. These findings have implications for potential prevention/early treatment of the neurodegenerative synucleinopathies.

**Risk of disease**

As this is the largest study ever performed in iRBD, it has potentially the most precise estimates of phenoconversion rates. Overall, we found phenoconversion rates of 6.25% per year. This is broadly similar, although slightly lower than some previous estimates, including that of the only previous multicentre study (which found an 8% annual conversion; Postuma et al., 2015d). The reason for this
slightly lower estimate is unclear. One explanation could have been secular change; as a disease becomes increasingly recognized, milder/earlier cases (with lower conversion rates) come to attention. However, we found no clear evidence for this; those diagnosed after 2010 had a 19.2% 3-year risk of disease, compared to 16.9% among those before. It could also be possible that newer centres in the RBD Study Group might have different (i.e. more permissive) diagnostic procedures, which would imply increasing proportions of patients without true synucleinopathy. However, centres who participated in the original multicentre study did not have a higher risk than those without (e.g. original centres’ 3-year risk = 16.8%, versus 21.6% for new centres). We did note that annualized disease risk appeared to be lower from Years 0 to 2 than for subsequent years. This may indicate a potential selection bias; if examiners were reluctant to recruit patients who appeared on the threshold of parkinsonism or dementia, risk would be systematically underestimated (since patients would have to first develop mild signs, then full disease).

Another potentially key factor for phenoconversion may be the frequency and intensity of follow-up. Many patients do not recognize symptoms of parkinsonism/cognitive impairment, and are diagnosed only on in-person systematic examination. A striking illustration of the importance of follow-up intensity is the Montreal experience. In their 2009 report, which included patients followed clinically/ad hoc, conversion risk at 5 years was 18% (Postuma et al., 2009). However, 6 years later, a study from the same centre, this time concentrating exclusively upon patients followed systematically by a movement disorders specialist and neuropsychologist, found a 5-year risk of 47% (Postuma et al., 2015c). Moreover, we may see evidence of this in our cohort, as conversion estimates were higher when they were calculated starting from the first date of intensive in-person examination of Parkinson’s disease/dementia risk factors (olfaction, UPDRS, cognitive exam, etc). For example, if conversion risk is tracked from performance of the first UPDRS Part III neurologist examination (a potential sign that more intensive follow-up has commenced), the estimated annual risk of conversion rises from 6.3% to 7.1%; see Table 4 for the potential effects of this on observed versus estimated sample size calculations. This might imply that a clinical trial with intensive...
periodic evaluations may find a higher conversion risk than observed in this study. Regardless of the conversion rate, it is clear that the large majority of idiopathic RBD patients in fact have prodromal synucleinopathy. So, while the term ‘idiopathic’ RBD is used here, we recognize that few patients are truly ‘idiopathic’ in the original sense of the term (i.e. unclear cause), and other terms such as ‘clinically isolated’ RBD may be more appropriate (Holg et al., 2018).

**Predictive markers**

Although comparisons of hazard ratios across different predictors should be made with caution (because centres measured different variables), it nonetheless suggests numerous findings of interest. When analysed as a binary diagnostic test, there was no clear advantage of DAT-SPECT over either the UPDRS or quantitative motor testing (note that both DAT-SPECT and quantitative motor tests were defined by each centre as normal/abnormal with no harmonization procedures; harmonization might increase the hazard ratio). Note that this finding may be unique to iRBD patients, who have an extremely high prevalence of underlying synucleinopathy; in the general population, non-specific causes of motor slowing on quantitative motor tests (e.g. arthritis) may influence estimates more (Keézer et al., 2016; Jennings et al., 2017). Regardless, these quantitative motor tests were simple office-based tests that required <5 min to administer. Clearly these are strong candidates for selecting patients for future neuroprotective trials, and could even obviate the need for sophisticated imaging techniques if simpler trial design is required. This finding illustrates both the need to improve imaging techniques for prodromal disease and the considerable future potential for more precise quantitative motor markers (e.g. wearable or smartphone-based sensors).

It is not surprising that the highest hazard ratios were for motor and cognitive measures, since these are the primary means by which parkinsonism and dementia are defined; however, the high performance of olfactory testing as a predictor is notable, as it is also easily tested in office settings. Finally, no test appeared to be able to ‘rule out’ phenoconversion; many of those with normal testing still went on to develop parkinsonism and dementia. For example, the highest negative predictive value was seen for the MDS prodromal criteria, but even among those negative for criteria, 5% phenoconverted at 3 years, 13% at 5 years, and 27% at 8 years (note that analysis is at baseline only, and presumably many of these patients would have developed abnormal markers before phenoconversion).

**Dementia-first versus parkinsonism-first**

The comparison between dementia-first and parkinsonism-first phenoconvertors was notable for the similarity in predictive value between markers. Motor variables were highly predictive of dementia as well as parkinsonism (and for quantitative motor assessment, even more predictive of dementia than parkinsonism). This finding is consistent with previous studies which documented a longer/slower-progressing motor prodromal interval in dementia-first than parkinsonism-first convertors (Postuma et al., 2012); if their motor prodromal interval is longer in prodromal dementia patients, they would be more likely to be abnormal on a cross-sectional test. Overall, the only clear differentiating variable between dementia and parkinsonism was cognition itself. It is unclear whether the conversion to dementia versus parkinsonism first is related to a different ‘top-down’ synuclein spread upwards to cortex before the substantia nigra (Adler and Beach, 2016), or to effects of co-morbid pathology [i.e. if a person with RBD has co-morbid amyloid cortical pathology, even modest cortical deposition of synuclein could trigger rapid cortical neurodegeneration resulting in a dementia-first phenotype (Chetelat et al., 2013)].

**Sample size**

We calculated the sample size requirements for a definitive neuroprotective trial, using phenoconversion as a categorical endpoint. Overall, sample sizes for a 2-year trial with HR = 0.5 ranged from 150 to 360 patients per group. In general, stratification strategies could decrease sample sizes, at the cost of reduced generalizability and less efficient recruitment. Of the selection strategies, the two most efficient appeared to be olfaction, which reduced sample size by 28.5% while retaining 67% of the sample as potential trial candidates, and the MDS prodromal criteria, which reduced sample size by 17.8% while retaining 77% of the sample. Of course, exact sample size calculations will depend on the specifics of a clinical trial; nevertheless, the fact that 24 centres combined to produce these estimates can provide some confidence for trial planners that sample sizes will be representative of the global experience. Notably, the total sample size for a future neuroprotective trial is less than the number of participants who were recruited to this study. So, it appears that a complete trial-ready population already exists in the centres of the International RBD Study Group.

**Limitations and strengths**

Some limitations of this study should be pointed out. First, this study is an amalgam of the research experience of 24 different centres; there was not a single protocol for testing predictors of disease, and protocols differed greatly between centres in terms of depth, follow-up intensity, predictors assessed, and methods/cut-offs for assessing them. Therefore, the predictive data will not be fully comparable to a single clinical trial setting, which would have a single testing protocol. Second, protocols for recruiting MCI varied; 23 of 24 centres recruited patients at baseline with MCI, but the largest centre (Barcelona) did not. There is no perfect way to harmonize these completely; for the primary analysis we elected to allow the Barcelona group to define disease conversion as de novo MCI, to prevent underestimation of...
disease risk (i.e. if patients with MCI were systematically excluded at baseline, then patients developing dementia would have to cycle from normal cognition through MCI to dementia, artificially prolonging disease-free time). However, if conversion from normal cognition to MCI were faster than from MCI to dementia/parkinsonism, disease risk might be overestimated. Regardless, excluding Barcelona data had almost no effect on risk estimates (median conversion time = 8.01 years with and 8.00 without). Third, hazard ratio comparisons between the different markers should be made with caution, as different centres (with potentially different conversion rates) tested different markers using different techniques (note that the results are adjusted for centre, which helps mitigate centre effects). Fourth, the amplitude of the hazard ratio observed in this study should not be extrapolated to the general population. When using RBD patients, the baseline risk of disease is so high that ceiling effects on hazard ratios occur [for illustration of this effect, see supplemental methods of Berg et al. (2015)]. Similarly, the effect of very long latency prodromal markers (e.g. autonomic dysfunction, olfaction, substantia nigra ultrasound) may be masked by floor effects; if a marker preceded RBD in almost all cases, and almost all RBD patients have prodromal synucleinopathy, there would be little apparent predictive value of that marker in this population. Fifth, RBD in Parkinson’s disease marks a ‘diffuse-malignant’ subtype of Parkinson’s disease (Fereshtehnejad et al., 2015), implying that our hazard ratio findings will not completely generalize to those Parkinson’s disease and dementia with Lewy bodies cases who have no RBD. Sixth, note that our markers were tested at baseline only; repeated marker testing would allow assessment of evolution of prodromal markers over time. Seventh, although sample size is large in this trial, some markers were assessed by only a few centres, and so their corresponding confidence intervals can be wide. Eighth, the final neurodegenerative disease diagnosis of all patients in this study was clinical, according to best impression of the treating neurologist; it is likely that some patients diagnosed with Parkinson’s disease will eventually be discovered to have MSA, and vice versa. Finally, the number of patients with very long duration follow-up remains limited (e.g. 28 still-disease-free patients have been followed for >12 years); therefore, we cannot determine whether disease risk changes over very long disease durations.

In conclusion, we confirmed a high risk of phenoconversion to overt neurodegenerative disease in RBD and found numerous predictors of phenoconversion. As new disease-modifying treatments are being developed for neurodegenerative synucleinopathies, RBD patients are ideal candidates for neuroprotective trials.

Funding

This study was funded by a grant from the Canadian Institute of Health Research (Grant #286641) the Weston Foundation (grant #N/A) and the Fonds de la Recherche en Santé Quebec (#28915). The funding sources had no role in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at Brain online.

References


