Forecasts for the Attainment of Major Research Milestones in Parkinson's Disease

Patrick Bodilly Kane, PhD[1], Daniel M Benjamin, PhD[2], Roger A Barker, PhD[3], Anthony E Lang, MD[4], Todd Sherer, PhD[5], Jonathan Kimmelman, PhD[1*]

1. Biomedical Ethics Unit, STREAM Research Group; McGill University, 3647 Peel St, Montreal, QC, H3A 1X1; Phone: (514) 398-3306; Fax: (514) 393-8349; Email: jonathan.kimmelman@mcgill.ca.

2. University of Southern California, Information Sciences Institute, 4676 Admiralty Way, Suite 1001, Marina del Rey, CA 90292. Email: benjamin@isi.edu.

3. John van Geest Centre for Brain Repair, WT/MRC Cambridge Stem Cell Institute, Department of Clinical Neuroscience, University of Cambridge, Cambridge CB20PY, UK. Email: rab46@cam.ac.uk.

4. Edmond J. Safra Program in Parkinson's Disease and the Morton and Gloria Shulman Movement Disorders Clinic, Toronto Western Hospital. Email: Anthony.Lang@uhnresearch.ca.

5. The Michael J. Fox Foundation for Parkinson's Research; 111 West 33rd Street, 10th Floor, New York, New York 10120; Phone: (212) 509-0995; Fax: (212) 509 2390; Email: tsherer@michaeljfox.org

* Corresponding author

ABSTRACT

Background: Projections about when research milestones will be attained are often of interest to patients and can help inform decisions about research funding and health system planning.

Objective: To collect aggregated expert forecasts on the attainment of 11 major research milestones in Parkinson's disease (PD).

Methods: Experts were asked to provide predictions about the attainment of 11 milestones in PD research in an online survey. PD experts were identified from: 1) The Michael J. Fox Foundation for Parkinson's Research data base, 2) doctors specializing in PD at the ranked neurology centers in the US and Canada, and 3) corresponding authors of articles on PD in top medical journals. Judgments were aggregated using coherence weighting. We tested the relationship between demographic variables and individual judgments using a linear regression.

Results: 249 PD experts completed the survey. In the aggregate, experts believed that new treatments like gene therapy for monogenic PD, immunotherapy and cell therapy had 56.1%, 59.7%, and 66.6% probability, respectively of progressing in the clinical approval process within the next 10 years. Milestones involving existing management approaches, like the approval of a deep brain stimulation device or a body worn sensor had 78.4% and 82.2% probability of occurring within the next 10 years. Demographic factors were unable to explain deviations from the aggregate forecast ($R^2=0.029$).

Conclusions: Aggregated expert opinion suggests that milestones for the advancement of new treatment options for PD are still many years away. However, other improvements in PD diagnosis and management are believed to be near at hand.

Introduction

Parkinson's disease (PD) research and development is progressing along many fronts, including precision medicine, experimental new therapies, and body worn sensors for diagnosis and monitoring. Though major advances in disease modifying therapy have yet to arrive, novel therapeutic strategies like deep brain stimulation or better levodopa infusion have improved the symptomatic management of PD.

Because PD has a protracted course, patients and physicians are often keenly interested in expert impressions of existing research, which can help estimate when new treatments and management strategies are expected to emerge and be trialed. Realistic projections of the attainment of major research milestones can help physicians and patients calibrate their expectations. They can also help with research and health system planning. For example, knowing when novel interventions are likely to mature can help healthcare systems determine when to build capacity for their deployment. Knowing which therapeutic strategies are likely to mature first can help funding bodies set priorities or issue calls that are relevant to that milestone.

Forecasting scientific breakthroughs is often not amenable to computational approaches, in part due to the diversity of variables that drive scientific progress.[1] In what follows, we used a "wisdom of the crowd" approach for aggregating expert forecasts about the timing of eleven major research milestones for the future management and diagnosis of PD. Such approaches have been shown to offer greater predictive accuracy than individual judgments by reducing the effect of random variation between expert judgments, while pooling the judgments of individuals who have widely varying knowledge.[2–4] Previously, wisdom of crowds approaches have been

used in medicine to improve predictive accuracy in areas like prognosis[5], diagnosis[4,6], and emergence of diseases.[7–9] Our approach used a method known as coherence weighting, where the aggregation of forecasts is weighted based on the extent to which the forecasts are probabilistically coherent. This approach has demonstrated effectiveness in improving the accuracy of lay forecasts in other domains, where those who provided more coherent forecasts also provided more accurate ones.[10]

Methods

PD experts were recruited from two sources: 1) the Michael J. Fox Foundation for Parkinson's Research database and 2) an independent sample constructed by identifying doctors specializing in PD at the 25 top ranked neurology departments in the US (as rated by US News and World Report), the top 3 largest hospital systems in Canada, and by identifying corresponding authors of articles on PD from the last 5 years in *BMJ*, *Lancet*, *PLoS Med*, *NEJM*, *JAMA*, *Annals of Neurology, JAMA Neurology*, *Lancet Neurology*, *Neurology* and *Movement Disorders*. Experts were solicited for survey participation three times.

Our survey sought forecasts for 11 major milestones in PD research (this was the largest number of forecast queries that could fit in a 30-minute survey). Milestones were generated by our three co-authors with PD expertise (RB, AL, TS) based on their perceptions of what would constitute a significant improvement on the current state of the field. We then helped select milestones that were objectively verifiable, diverse, and thought to be of interest to the patient community. Briefly, each PD expert co-author generated 10-15 potential milestones, and then rated the

milestones created by the other co-authors on their importance and verifiability. We selected the highest rated milestones while ensuring that some selected milestones were likely to occur earlier while others were likely to occur later and covered a range of event types including FDA approval and trial launch. Particular milestone events for new treatments (e.g. FDA approval vs. launch of a trial vs. clinical practice guideline recommendations) were selected in order to enable capture of near term as well as late-occurring advances. The survey was created on Qualtrics. Experts provided the probability of milestone attainment in three time bins spanning the next 10 years (within the next 2 years, within the next 2-6 years and within the next 6-10 years), and separately the probability of milestone nonattainment within 10 years. In order to assess probabilistic coherence for the purpose of coherence weighting, we allowed experts to enter probabilities that did not sum to 100% across time bins. A graphical depiction of our survey process is provided in **Figure 1**. Milestone questions are provided in **Table 1**. Milestones were presented in random order.

Our survey contained five additional components. First, for milestones entailing launch or completion of a trial, we asked experts to predict the probability the trial would be positive on its primary endpoint. Second, we collected forecasts of whether there would be an FDA approval within the next 10 years for therapies in the following areas: gene therapy, repurposed small molecule therapy, novel small molecule therapy, stem cell therapy or a therapy not falling within the preceding categories. Note that, unlike the 11 milestone questions, these questions were not worded with sufficient precision to enable forecast verification. We additionally asked participants to rate their familiarity with gene therapy, precision medicine, alpha-synuclein targeted treatment and deep brain stimulation on a seven point Likert scale from extremely

unfamiliar to extremely familiar. Third, we asked participants to rate the same four kinds of

treatments on their clinical promise on a nine point Likert scale from minimal promise to

maximal promise. None of the above questions were asked using a coherence weighting format.

Last, we collected the following demographic items from experts: sex, age, education and

number of clinical trials participated in. H-indices of all participants were obtained using Scopus.

Our survey received approval by McGill IRB; experts provided consent online.

Aggregated forecasts were produced in two stages. First, the raw forecasts were all made

coherent by ensuring that the forecasts for each time bin jointly summed to 100%. We did this

by taking each set of incoherent forecasts, and finding the coherence set of forecasts that was

closest to them as measured by the Euclidean distance. For example, if someone provided

forecasts of 30%, 40%, 40%, 30% for the four time bins their forecasts would be incoherent,

because they sum to more than 100% percent. These forecasts would be coherentized by finding

the closest set of coherent forecasts, in this case 20%, 30%, 30%, 20%.   Once the forecasts were

coherentized, we averaged them together to produce an aggregate forecast. This average was

weighted by how incoherent each forecast was originally, that is how far from 100% the sum

was, with lower weight given to more incoherent forecasts.[10] Post-hoc, we tested whether the

forecasts for North American experts were different from the forecasts for non-North American

experts for the two milestones dealing with FDA approvals using a Kolmogorov-Smirnov test for

equality of distributions.

As an exploratory analysis we tested for differences in the deviation of forecasts from the coherence weighted mean as a function of the demographic characteristics of experts. Under the assumption that the coherence weighted mean is the best obtainable prediction for each milestone, this dependent variable serves as a proxy for forecast accuracy. To simplify the analysis, we first dichotomized forecasts by summing each individual's forecasts for each milestone for the first three time bins to create a forecast of the probability of the milestone occurring in the next 10 years. We performed a similar dichotomization for the coherence weighted means, and then took the absolute value of their difference. We subjected this dependent variable to a linear regression on age, H-index, number of clinical trials participated in, an indicator variable for gender, a set of indicator variables for degrees held and incoherence, along with milestone controls. Based on a preliminary exploration of the data, we also developed a model where we used a log transformation on H-index and number of trials participated in to deal with some of the skew in these variables. When we compared models we found that the fit was not notably different (R^2=0.029 versus R^2=0.029), so we report the results for the untransformed model only.

Data Availability Statement: CSV files containing both raw and coherentized forecasts as well as demographic information will be made available by request.

Results

Characteristics of Expert Participants

249 experts in PD completed our survey; 87 were recruited through the MJFF database (of 2092 contacted) while 162 were recruited from our independent sample (of 811 contacted). Median age of respondents was 48 (range 24-86); 31% of respondents were female; degrees held by our respondents were MD (24%), MD/PhD (27%) PhD (36%), and Master or less (13%); median H-index of respondents was 22.5 (standard deviation 24.27, range 1-192). Median H-index of the population from which our independent sample was drawn was 25 (standard deviation 26.95, range 1-194). A Kolmogorov-Smirnov test for equality of distributions did not reject the null hypothesis that our sample and the population had the same distribution of H-indices (D=0.81, p=0.21). Experts were located in North America (46%, with 35% total from the USA), Europe (35%), Asia (7%), Oceania (3%), and Africa and South America (1% each). We were unable to obtain location information for 5% of our respondents.

Forecasts on Milestone Attainment

The median incoherence was 14.5%, (26.4% standard deviation, range 0%-127%), meaning that forecasts of half of experts summed to either less than 85.5% or more than 114.5%. Forecasts about the timing for milestone attainment are depicted in **Figure 2** (the distribution of individual forecasts for each milestone can be found in the Supplementary materials). Three of the four milestones judged to be least likely to be attained in the next 10 years concerned disease modifying therapies, while those judged most likely to occur are largely refinements in existing therapies. The aggregated predictions of experts for trials being positive on their primary endpoint was 44.5% for the trial specifying eligibility based on GBA mutational status trial (Precision Medicine Therapy), 39.4% for the pluripotent stem cell trial (Cell Therapy), 47.0% for the trial integrating an alpha synuclein imaging agent (Imaging), and 40.5% for the non-

cholinesterase inhibiting drug trial (Treatment for PD-MCI). **Table 2** contains the mean and

standard deviation for each time bin for the FDA related milestones for North American and

non-North American experts, as well as the results of a Kolmogorov-Smirnov test for equality of

distribution comparing the distributions of predictions. None of the tests were significant

indicating there was no difference in the distribution of these predictions between North

American and non-North American samples.

Judgments on Treatment Categories and Clinical Promise

The mean prediction for FDA approval in the next 10 years for the different treatment categories

are displayed in **Table 3**. The mean familiarity and clinical promise ratings with different

treatment categories are displayed in **Table 4.**

Relationship Between Expert Characteristics and Forecasts

For our exploratory analysis, we evaluated 2348 forecasts out of 2733 available forecasts (for 6

experts, we were missing all demographic information; for another, 35 we were missing data on

one or more covariates in the regression model). The model accounted for very little of the

variation in the data (R^2=0.029). The coefficients are listed in **Table 5**. The coefficients on age,

clinical trials participated in and H-index were not significantly different from zero (t=-0.09,

p=0.93; t=-0.76, p=0.45; t=0.27, p=0.79). The factor describing education significantly

improved model fit ($F(3, 2348)$=2.85, p=0.04); those with MDs, MD-PhDs and PhDs all

provided predictions that were further from the coherence weighted mean than those with only a

Masters degree. The factor describing gender also significantly improved model fit ($F(2,$

$2348)$=3.84, p=0.02), with male participants providing predictions further from the coherence

11

weighted mean than female participants. Overall, this regression analysis suggests that there is a great deal of unexplained variation in the data, and the only variables with any predictive power appear to be education and gender.

Discussion

We synthesized a large and diverse sample of expert opinions about the timing of attainment of major research milestones in PD. With the exception of precision medicine approaches, experts believed that advances in new treatment modalities are likely still some years away, viewing breakthroughs in disease modifying treatments as having a 30-40% chance of not occurring at all within ten years. Experts were most pessimistic about FDA approval of Parkinson's monogenic gene therapy relative to other disease modifying therapies, though this may have more to do with the milestone referring to an FDA approval as opposed to a clinical trial. When asked about broad treatment categories rather than more specific milestones, experts rated the probability of the FDA approving a new treatment within 10 years as less than 50% for all categories. Doubts about near term breakthroughs in disease modification were echoed by forecasts that primary endpoints in trials are more likely to be non-positive than positive, and the fact that experts generally predicted a less than 50% probability of an FDA approval of a new PD therapy across five different treatment platforms. However, on average experts rated the clinical promise of all treatment categories we asked about as above the mid point of our scale, indicating that they believe these lines of research are worth pursuing even though they are less likely to produce concrete gains in the next 10 years.

Experts did, however, believe improvements for existing therapy and improvements in diagnostic techniques are likely in the near future. For example, experts predicted a 2 in 3 chance that a repositioned drug will demonstrate disease-modifying activity in patients within ten years. Experts anticipated that clinical practice guidelines are more likely than not to endorse the use of body worn sensors for PD diagnosis within the next six years.

Our regression analysis of the relationship between expert characteristics and forecasts suggest that simple demographic characteristics shed very little light on "accuracy". The only significant contributions to model fit were that those individuals with a Master's degree or less tended to make predictions closest to the coherence weighted mean, and that men tended to make predictions further from the coherence weighted mean than women. However, both effects were small. If the coherence weighted mean is in fact an accurate assessment of the true probabilities, this analysis implies that looking at demographic variables, even those that supposedly track expertise, may not be the optimal experts with the best predictive abilities. This would suggest that when decisions like funding or priority setting hinge on assessing the timing of scientific advances, granting agencies or policy-makers may be better off randomly recruiting advisors from a list of established experts rather than seeking out rarefied (and often expensive) expertise.

Our study has limitations. The first concerns milestone sampling. The wording of our survey questions was very specific and may not have captured the promise of broader milestone categories. For example, our gene therapy milestone excluded non-viral vector therapies like anti-sense oligonucleotides, and only concerned treatments for monogenic PD. As such, answers to this question should not be viewed as proxies for all gene therapy approaches being tested.

The use of specific milestone questions reflected the scientific imperative in forecast research of ensuring that each question is clear, unambiguous, and verifiable. The inclusion of more general questions might mitigate this limitation. Related to this, many milestone questions concerned outcomes in the U.S. drug approval process. As a consequence, forecasts reflected beliefs about both clinical promise and the pacing and standards used for the approval process itself. To the latter point, non-North American based experts may have found the task of forecasting regulatory attainment milestones more difficult, though their forecasts on the FDA related milestones appear similar overall to the those given by the North American based experts. The second set of limitations concern our expert sampling. Despite coherence weighting and selective eligibility criteria, forecasts could have been affected by response bias. Even so, our sample was largely composed of experts who had records of research productivity and who were affiliated with top neurology programs. Our demographic analysis of nonresponders did not suggest striking biases in our sample. The third set of limitations concern the predictions themselves. It remains to be seen whether the wisdom of the crowd approach we used will provide an accurate assessment of timelines. Further, the very availability of predictions reported in our study could change the probability of milestones being achieved. A last set of limitations concerns secondary analyses. A more comprehensive collection of demographic and cognitive features of experts might reveal other important factors that relate to forecast skill. Also, this analysis was conducted after the data had been analyzed for our primary objectives; all p-values should be understood as hypothesis generating.

That experienced experts did not converge more rapidly on aggregated expert opinion should not be interpreted as questioning the value of expertise. More experienced experts are likely to

contribute to funding and policy decisions in other important ways, including the identification of factors that need to be considered when making forecasts.[11] Our survey nevertheless provides a meaningful synthesis of state-of-the art expert opinion on the expected timing of several major breakthroughs in Parkinson's disease research. Many patients and caregivers show intense interest in learning about emerging new treatments; their expectations are often buffeted by hyperbolic claims in the press, on the internet or from pharmaceutical companies with a vested interest in particular therapies. Our forecasts provide a more objective representation of how expert communities interpret available evidence about when major advances will occur. They also provide healthcare system planners with an appraisal of the level of optimism about the availability of new therapies, diagnostics, and research techniques in the coming decade. Ultimately, the approach we employed of soliciting expert forecasts, and weighting them using coherence, can also be of use in helping funders, like disease charities or pharmaceutical companies, access accurate expert judgments of where to invest their resources. The approach of elicitation and aggregation is more likely to avoid many of the biases that accrue due to psycho-social dynamics that emerge with committees[12] or other expert elicitation platforms.[13]

Author Contributions

1) Research project:

      A. Conception: DMB, RAB, AL, TS, JK

      B. Organization: DMB, JK

      C. Execution: PBK, DMB, JK

2) Statistical Analysis:

      A. Design: PBK, DMB

B. Execution: PBK
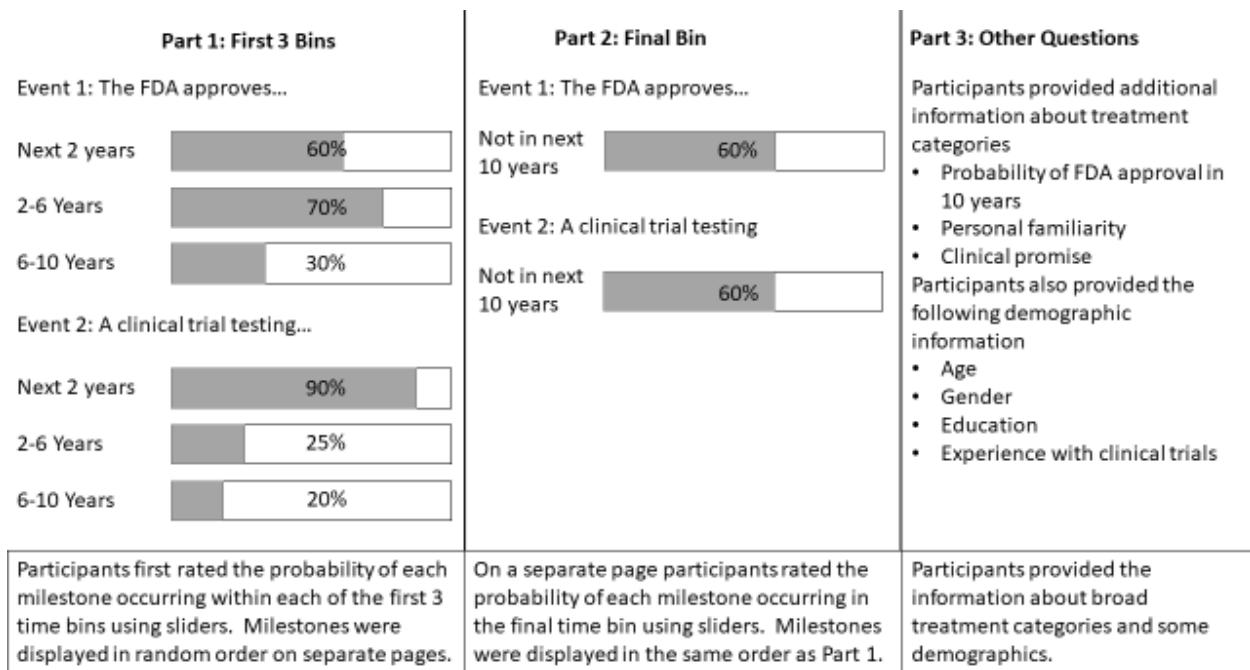
C. Review and Critique PBK, DMB, JK

3) Manuscript:

A. Writing of the first draft: PBK, JK

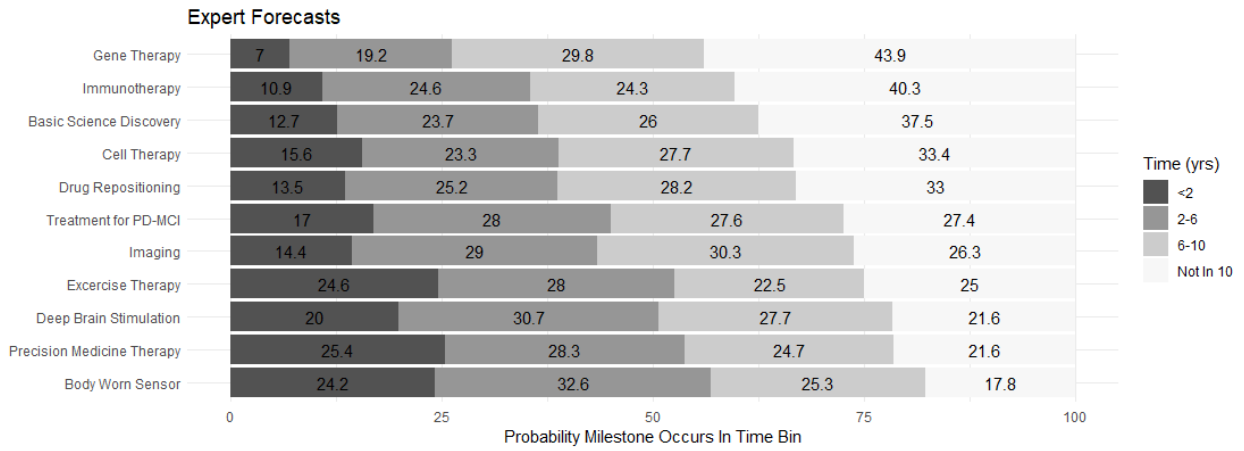B. Review and Critique: DMB, RAB, AL, TS

References

[1]   Clauset A, Larremore DB, Sinatra R (2017) Data-driven predictions in the science of science. *Science* **355**, 477–480.

[2]   Surowiecki J (2004) *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*, Doubleday & Co, New York, NY, US.

[3]   Armstrong JS (2001) Combining forecasts. In *Principles of forecasting: a handbook for researchers and practitioners*, Armstrong JS, ed. Kluwer Academic Publishing, Norwell, MA, pp. 417–439.

[4]   Kurvers RHJM, Herzog SM, Hertwig R, Krause J, Carney PA, Bogart A, Argenziano G, Zalaudek I, Wolf M (2016) Boosting medical diagnostics by pooling independent judgments. *PNAS* **113**, 8777–8782.

[5]   Kattan MW, O'Rourke C, Yu C, Chagin K (2016) The Wisdom of Crowds of Doctors: Their Average Predictions Outperform Their Individual Ones. *Med Decis Making* **36**, 536–540.

[6]   Kurvers RHJM, Krause J, Argenziano G, Zalaudek I, Wolf M (2015) Detection Accuracy of Collective Intelligence Assessments for Skin Cancer Diagnosis. *JAMA Dermatol* **151**, 1346–1353.

[7]   Li EY, Tung C-Y, Chang S-H (2016) The wisdom of crowds in action: Forecasting epidemic diseases with a web-based prediction market system. *Int J Med Inform* **92**, 35–43.

[8]   Neumann GR, Nelson FD, Polgreen PM (2006) Using Prediction Markets to Forecast Trends in Infectious Diseases. *Microbe Magazine* **1**, 459–465.

[9]   Tung C, Chou T, Lin J (2015) Using prediction markets of market scoring rule to forecast infectious diseases: a case study in Taiwan. *BMC Public Health* **15**, 766.

[10]   Karvetski CW, Olson KC, Mandel DR, Twardy CR (2013) Probabilistic Coherence Weighting for Optimizing Expert Forecasts. *Decision Analysis* **10**, 305–326.

[11]   Tetlock PE, Gardner D (2015) *Superforecasting: The art and science of prediction.*, Crown Publishers/Random House, New York, NY, US.

[12]   Sunstein CR (2006) *Infotopia: How Many Minds Produce Knowledge*, Oxford University Press, Inc., USA.

[13]   Morgan MG (2014) Use (and abuse) of expert elicitation in support of decision making for public policy. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7176–7184.

Part 1: First 3 Bins	Part 2: Final Bin	Part 3: Other Questions
Event 1: The FDA approves... Next 2 years — 60% 2-6 Years — 70% 6-10 Years — 30% Event 2: A clinical trial testing... Next 2 years — 90% 2-6 Years — 25% 6-10 Years — 20%	Event 1: The FDA approves... Not in next 10 years — 60% Event 2: A clinical trial testing Not in next 10 years — 60%	Participants provided additional information about treatment categories • Probability of FDA approval in 10 years • Personal familiarity • Clinical promise Participants also provided the following demographic information • Age • Gender • Education • Experience with clinical trials
Participants first rated the probability of each milestone occurring within each of the first 3 time bins using sliders. Milestones were displayed in random order on separate pages.	On a separate page participants rated the probability of each milestone occurring in the final time bin using sliders. Milestones were displayed in the same order as Part 1.	Participants provided the information about broad treatment categories and some demographics.

Figure 1: A simplified depiction of the interface and flow of our study.

Expert Forecasts

Milestone	<2	2-6	6-10	Not In 10
Gene Therapy	7	19.2	29.8	43.9
Immunotherapy	10.9	24.6	24.3	40.3
Basic Science Discovery	12.7	23.7	26	37.5
Cell Therapy	15.6	23.3	27.7	33.4
Drug Repositioning	13.5	25.2	28.2	33
Treatment for PD-MCI	17	28	27.6	27.4
Imaging	14.4	29	30.3	26.3
Excercise Therapy	24.6	28	22.5	25
Deep Brain Stimulation	20	30.7	27.7	21.6
Precision Medicine Therapy	25.4	28.3	24.7	21.6
Body Worn Sensor	24.2	32.6	25.3	17.8

Probability Milestone Occurs In Time Bin

Figure 2: Coherence weighted mean forecasts for each milestone, arranged in descending order of probability of nonattainment within 10 years. Each bin contains the coherence weighted mean predicted probability of the milestone occurring in that bin.

Topic	Event	Full description
Gene Therapy	FDA approval	The FDA approves a gene therapy directed at a monogenetic cause of PD such as LRRK2, GBA or parkin for treatment of Parkinson's disease.
Immuno-therapy	Trial results	A rigorous phase 2 or phase 3 clinical trial testing an alpha-synuclein based immunotherapy for Parkinson's disease reports a positive outcome on a primary efficacy endpoint.
Basic science discovery	Awarded *Science* magazine's Breakthrough of the Year	*Science* magazine awards "Breakthrough of the Year" to a molecule, process, cell, or discovery that is expressly described, in the accompanying Science article, as implicated in Parkinson's disease pathogenesis or possible treatment.
Cell Therapy*	Trial initiation	The launch of a rigorous phase 2 or phase 3 clinical trial involving implantation of Parkinson's patients with dopaminergic cells derived from pluripotent stem cells.
Drug repositioning	Trial results	A rigorous phase 3 clinical trial utilizing a repositioned medication and aimed at slowing the progression of Parkinson's symptoms reports a positive outcome on a primary efficacy endpoint.
Treatment for PD-MCI*	Trial initiation	Launch of a rigorous phase 3 clinical trial testing a novel, non-cholinesterase inhibiting drug in the treatment of PD-MCI.
Imaging*	Trial initiation	A selective alpha synuclein imaging agent is integrated into a rigorous Parkinson's interventional clinical trial.
Exercise therapy	Trial results	A rigorous phase 2 or phase 3 clinical trial testing the effect of exercises, physical activity, or physical therapy on Parkinson's progression reports a positive outcome on a primary efficacy endpoint.
Deep Brain Stimulation	FDA approval	FDA approval of the first closed-loop deep brain stimulation device for the management of Parkinson's disease.
Precision Medicine Therapy*	Trial Enrolment	A rigorous phase 2 or phase 3 clinical trial in Parkinson's disease that specifies eligibility based on GBA mutational status successfully enrolls at least 80 subjects.
Body worn sensors	Clinical practice guideline recommendation	An algorithm derived from a body worn sensor is accepted by the International Parkinson and Movement Disorder Society as a valid measure of Parkinson's disease symptoms.

Table 1: List of milestones used in our survey. Additional clarifying details for some of the milestones were listed as footnotes (see Supplementary Materials). Milestones with asterisks included an additional question asking experts to forecast the probability the given trial would report a positive outcome on its primary endpoints.

Milestone	Time Bin	North American Mean (SD)	Non-North American Mean (SD)	P-value
Gene Therapy	Next 2 years	6.5% (10.9%)	8.0% (9.3%)	0.057
	2-6 years	18.9% (15.7%)	20.9% (14.7%)	0.45
	6-10 years	31.8% (19.2%)	29.9% (19.0%)	0.71
	Not in next 10 years	42.9% (27.4%)	41.1% (28.1%)	0.78
Closed Loop Device	Next 2 years	21.9% (24.0%)	17.0% (20.6%)	0.23
	2-6 years	31.8% (18.9%)	31.2% (17.7%)	0.77
	6-10 years	28.4% (22.8%)	31.2% (20.3%)	0.12
	Not in next 10 years	17.9% (22.1%)	20.6% (22.2%)	0.20

Table 1: Means and standard deviations of the forecasts for each time bin for the two milestones dealing with the FDA for the North American and non-North American experts, as well as the p-values for a two sided Komogorov-Smirnov test of equality of distributions for the two samples of forecasts for each time bin.

Treatment Category	FDA Approval Prediction Mean (SD)
Gene Therapy	43.9% (28.6%)
Repurposed Small Molecule Therapy	46.5% (26.1%)
Novel Small Molecule Therapy	45.4% (27.8%)
Stem Cell Therapy	34.2% (24.6%)
Therapy Not Falling in the Preceding Categories	37.4% (28.4%)

Table 3: Mean and standard deviation of experts' predictions of the probability of an FDA approval of a treatment within 10 years in five different categories.

Treatment Categories	Familiarity Mean (SD)	Clinical Promise Mean (SD)
Gene Therapy	3.8 (1.9)	5.5 (2.1)
Precision Medicine	4.3 (1.6)	6.3 (1.8)
Alpha Synuclein	4.2 (1.8)	5.9 (2.0)
Deep Brain Stimulation	4.9 (1.7)	6.3 (1.6)

Table 4: Mean and standard deviations for expert Likert scale ratings of familiarity with different kinds of PD treatment. Note that familiarity ratings were performed on a 7-point scale while clinical promise ratings were performed on a 9 point scale.

Variable	Coefficient	P-Value
Intercept (Gene Therapy, Female, Masters or less)	18.78	<0.001
Immunotherapy	0.64	0.62
Basic Science Discovery	0.86	0.51
Cell Therapy	-1.18	0.36
Drug Repositioning	-1.30	0.32
Treatment for PD-MCI	-3.63	0.005
Imaging	-3.08	0.02
Exercise Therapy	-0.42	0.75
Deep Brain Stimulation	-5.03	<0.001
Precision Medicine Therapy	-5.21	<0.001
Body Worn Sensor	-5.82	<0.001
MD	3.70	0.02
MD-PhD	3.14	0.05
PhD	4.22	0.007
Gender Male	1.79	0.007
Gender Unspecified	-1.09	0.80
Age	-0.003	0.93
Clinical Trials	-0.006	0.44
H-index	0.004	0.79

Table 5: Coefficients for the regression of deviation from the coherence weighted mean on demographic variables. Coefficients are interpreted as the change in deviation associated with a one unit change in the associated variables.