DR. TEA LUND LAURSEN (Orcid ID : 0000-0003-2494-0526)

Article type : Special Article

The Prevalence of Wilson disease. An Update.

Thomas Damgaard Sandahl[1] (tds@clin.au.dk), Tea Lund Laursen[1] (tealaurs@rm.dk), Ditte Emilie Munk[1] (dittmu@rm.dk), Hendrik Vilstrup[1] (vilstrup@clin.au.dk), Karl Heinz Weiss[2] (karlheinz.weiss@med.uni-heidelberg.de) & Peter Ott[1] (peterott@rm.dk).

From:

[1] Department of Hepatology and Gastroenterology, Aarhus University Hospital, Denmark.

[2] Department of Internal Medicine IV, University Hospital Heidelberg, Heidelberg, Germany.

Keywords: Hepatolenticular Degeneration; epidemiology; incidence; allele frequency; penetrance.

Contact information

Corresponding author:

Peter Ott

Department of Hepatology and Gastroenterology

Aarhus University Hospital

99 Palle Juul-Jensens Boulevard

8200 Aarhus N, Denmark

Phone: +45 40 19 38 26

FAX: +45 78 46 28 60

E-mail: peterott@rm.dk

List of Abbreviations

q : Wilson disease mutation allele frequency.

Conflict of interest

All authors declare no potential conflicts of interests (financial, professional, or personal) that are relevant to the manuscript.

Financial Support

The work was supported by a grant from The Memorial Foundation of Manufacturer Vilhelm Pedersen & wife. The foundation played no role in the planning or any other phase of the study.

Abstract

Background and Aims: In 1984, Scheinberg and Sternlieb estimated the prevalence of Wilson disease to be 1:30,000 based on the limited available data. This suggested a large number of overlooked cases with potentially fatal consequences. The "Scheinberg-Sternlieb Estimate" is still widely used although more recent clinical and genetic studies of higher quality are now available. In the present study, we included these data to update the prevalence estimate.

Methods: A MEDLINE Ovid, Science Citation Index Expanded, and PubMed systematic search for all relevant studies on Wilson disease prevalence was conducted.

Results: In total, 59 studies (50 clinical and 9 population-based genetic) were included in the final analysis. We identified four recent clinical studies based on nationwide databases of

high quality, providing prevalence estimates from 1:29,000 to 1:40,000. Higher frequency populations do exist due to frequent first cousin marriages and/or a higher mutation frequency. When calculating prevalence from the incidence related to number of births, estimates were 1:40,000-1:50,000. Clinical screening studies including examination for Kayser-Fleisher Rings or ceruloplasmin did not improve these estimates due to insufficient sample size or selection biases. Population-based genetic studies in US and UK populations were not in disagreement with the clinically-based estimates. At the same time studies from France and Sardinia suggested that the genetic prevalence may be 3-4 times higher than the clinical disease prevalence. This raises the question whether the penetrance is indeed 100% as generally assumed.

Conclusion: The original prevalence estimate from 1984 of 1:30,000-1:50,000 still appears valid at least for USA, Europe and Asia. In some population-based studies, the genetic prevalence was 3-4 times higher than clinically based estimates. The question of penetrance needs further evaluation.

Wilson disease is an autosomal recessively inherited disease caused by mutations in the *ATP7B* gene. The resulting dysfunction of the ATP7B protein leads to impaired copper excretion into the bile and causes a combination of hepatic, neurologic, and psychiatric symptoms. While effective treatments are available the disease will take a disabling and fatal course if the diagnosis overlooked and treatment not initiated.

Although known to be rare, the prevalence of the disease is a longstanding matter of interest in order to avoid overlooked cases. In 1968, Scheinberg and Sternlieb estimated the worldwide prevalence to be 1:200,000 based on the fragmented data available at the time (1). In 1984, they updated their estimate to 1:30,000 (30/million) based on 3 datasets: 1) The vital

statistics report from the US 1968-1978 (Wilson related cases 13.21/1,000,000), where they assumed that only half of the true Wilson cases were detected, 2) an East-German clinical population-based report on the prevalence (29/1,000,000), and 3) an advanced epidemiological analysis based on the frequency of first cousin parenthoods among 289 Japanese families with at least one family member with Wilson disease (33/1,000,000)(2). Scheinberg and Sternlieb regarded this estimate to be universal, except for small isolated populations with higher frequencies. At Hardy Weinberg equilibrium and 100% penetrance, this would correspond to a frequency of disease-causing mutated alleles of approximately 1:180 – 1:300 and a carrier frequency of 1:90 – 1:150. Even though the 1984 paper was based on limited and highly selected data, the resulting estimate of 1:30,000 is still widely used and regarded valid across ethnic differences and geographic zones. The 1984 paper suggested that a large number of patients were undiagnosed at that time with potentially fatal consequences. Even though the diagnostic awareness has increased and more cases are diagnosed today, it is uncertain to which extend Wilson patients are still overlooked.

During the last decades, a number of clinical epidemiological papers have been published on the prevalence of Wilson disease in different populations. As the public health statistics improves, estimates can now be based on still larger unselected populations from health care systems with increasing awareness for this diagnosis. This has improved data quality. In addition, current biotechnology allows for genetic screening for Wilson disease causing mutations in large populations, and such studies are emerging.

In the present paper, we reevaluate the current worldwide prevalence estimate of Wilson disease based on the currently available literature. We critically discuss how different methodologies may bias the estimates. We also examine what can be learned from the genetic

studies and how they will affect our understanding of the clinical epidemiology and the natural history of Wilson disease.

Methods

Literature search:

We searched MEDLINE Ovid (1946 to May 2018), Science Citation Index Expanded (Web of Science; 1900 to May 2018), and PubMed (Bethesda (MD): National Library of Medicine (US) 1966 to May 2018) using the search terms: Wilson's disease (all fields), Wilson disease (all fields), Wilson (in title) or Hepatolenticular Degeneration (all fields) or with subheading epidemiology (MeSH term) and Epidemiology (All fields or MeSH) or prevalence (All fields or MeSH) or incidence (all fields or MeSH). Additional studies were identified by a hand search of references of original articles or review articles and by personal communication with relevant investigators. Case reports were excluded except for countries where only case reports were available.

Methodological considerations

The paper is based on the following definitions and methodological considerations:

$$Disease\ prevalence = \frac{\text{Number of persons with Wilson disease}}{\text{Number of individuals in the population}}$$

The disease prevalence is affected by the number of patients who develop symptoms, the age at which it happens and the average life length of Wilson patients.

$$Crude\ prevalence = \frac{\text{Number of diagnosed cases}}{\text{Number of individuals in the population}}$$

The crude prevalence (or *clinical prevalence*) depends on the disease prevalence but is also affected by the number of overlooked cases (i.e. diagnostic awareness) and the diagnostic delay. Diagnostic delay of up to several years is still a considerable challenge (3-6). In this text, the term 'crude prevalence' is used pragmatically as the number of reported cases, so it will be affected also by diagnosed but unreported cases.

Wilson disease mutation allele frequency (q)

$$q = \frac{\text{Number of alleles with ATP7B mutations that are known to cause Wilson disease}}{\text{Number of ATP7B alleles in the population}}$$

At Hardy-Weinberg equilibrium, the risk that a person is homozygous or compound heterozygous for Wilson disease causing mutations in ATP7B is q^2, that of being a carrier is $2q(1-q)$ and that of having two wildtype alleles is $(1-q)^2$. The underlying assumptions: Infinite population size, random breeding, and absence of selection and genetic drift are often challenged. Worldwide, consanguinity, such as first cousin marriages, is an important deviation from these assumptions.

$$\textit{Genetic prevalence} = \frac{\text{Number of persons who are homozygous or compound heterozygous for known disease} - \text{causing mutations in the ATP7B gene}}{\text{Number of individuals in the population}}$$

The term "genetic prevalence" will be used because it is shorter than the more correct "prevalence of homozygocity or compound heterozygocity for known disease causing alleles in the ATB7B gene". At Hardy-Weinberg equilibrium, the genetic prevalence will equal q^2. Consanguinity will increase the genetic prevalence above q^2.

The disease prevalence will always be lower than the genetic prevalence because of factors such as age at presentation above zero, reduced life expectancy, delayed diagnosis, and overlooked cases. Because the average age of presentation is around 25 years, disease prevalence will maximal be ~0,7 times the genetic prevalence. This difference will increase if the penetrance is lower than 100%.

The use of the genetic prevalence to estimate the disease prevalence requires knowledge whether specific combinations of genetic variants cause Wilson disease. Since only a few mutations are characterized in detail e.g. H1069Q, R778L, and –441/–427del (7-11) , it is not always known if a genetic variant is disease-causing.

Penetrance

Penetrance is defined as *the risk that a person who is homozygous or compound heterozygous for known disease-causing mutations in the ATP7B gene will develop Wilson disease at some point during life*. For Wilson disease it is generally assumed to be 100%.

Statistics

Most of the reports on incidence or prevalence of Wilson disease do not include confidence limits. If not reported, they were calculated according the method described by Robert Newcombe, derived from the so called Wilson procedure (12).

Results

In total, 642 studies of Wilson disease were identified. We excluded in vitro studies, animal studies and studies with no apparent information about the prevalence of Wilson disease. We finally identified a total of 59 pertinent articles. Studies were categorized according to the applied methods. Fifty were categorized as clinical and 9 as population-based genetic studies.

Clinical studies

Isolated populations with high crude prevalence

Reports on isolated populations with very high risk of Wilson disease include a village in Crete, Greece (1:15 births) (13), Kalymnos, Greece (1:740) (14), and isolated mountain areas in Romania (1:1,130) (15), even though low sample size induce some uncertainty about the exact prevalence estimates. As expected, specific mutations dominate these populations (16-18) and high frequencies of these mutations and consanguinity both contribute.

Larger populations

A number of studies providing data for estimation of crude prevalence in individual countries were identified by the literature search (Supplementary table 1). The quality of these studies varies substantially likely due to overlooked and unreported cases. Many studies are from tertiary centers and may not include all the patients in the country. Most studies do not describe the origin of their patients, i.e. natives or immigrants. The description of the Austrian cohort (19) which has been maintained for a long time may be one the most complete with a crude prevalence of 1:39,800. However, even that study was not carried out with the intention to produce a prevalence estimate. Altogether, the results and quality of these studies vary too much to provide a valid estimate of the clinical prevalence.

Three recent large, nationwide, high quality populations studies bases on electronic records

are now available from France (20), Taiwan (21) and Hong-Kong (22). The crude prevalence estimates were 1:63,000, 1:55,000 and 1:40,000 respectively. However, the data clearly suggested overlooked cases. Thus, in the French study, the prevalence was higher in younger age groups and near larger cities and highest in the 20-29-year age group (1:37,000). In the Taiwan study, the crude prevalence increased from 1:90,000 in 2001 to 1:35,000 in 2011. The male/female ratio was 1.75/1 suggesting a number of undiagnosed cases among females, because the male/female ratio is 1(23). Since most Taiwan patients were diagnosed in neurological departments (21), hepatologic cases in which females predominate could have been overlooked. Thus, the best estimate for Taiwan is 1:29,100 as reported for males in 2011. In Hong-Kong, the prevalence increased from year 2000 and became stable after 2011 without sex differences and suggesting a crude prevalence of 1:40,000 to be realistic also in the future. Thus, the four best clinical studies provided estimates of the disease prevalence of 1:39,800 (19), 1:37,000(20), 1:29,000 (21), and 1:40,000 (22), respectively.

Some countries, like Israel, Croatia, Costa Rica and the Italian island Sardinia may have higher prevalence of Wilson disease, either related to consanguinity or higher mutation frequencies. There is no clear geographically east/west or north/south gradient (supplementary table 1).

Data from the Middle East, India, Pakistan, Africa and South America are so limited that prevalence estimates are not possible for these parts of the world (24, 25).

Prevalence related to number of births

Studies of the crude prevalence may severely underestimate the disease prevalence because of overlooked cases in the past. To overcome this bias, investigators have related the number of diagnosed cases to the number of births in a given period. Reports from such studies are available only from Europe and listed in Table 1. Estimates center around 1:40,000-1:50,000

except for a higher prevalence in Sardinia of 1:16,700. These numbers are not different from 1:29,000 – 1:40,000 derived from the three large high-quality studies on crude prevalence mentioned above (20-22). The Eurowilson collaboration reported incidence data from Croatia (2.44 WD/mio/3 years) and Austria (2.2 WD/mio/3 years) which may suggest somewhat higher prevalence, but these data were only published as an abstract (26).

Screening studies

Clinical or biochemical screening

To overcome the problem of overlooked cases, and the other methodological issues mentioned above, attempts have been made to screen for Wilson disease by biochemical or clinical methods, as listed in Table 2. These studies employed screening of larger populations for Kayser-Fleischer (KF)-rings (27), or ceruloplasmin in either whole dried blood (28-31) or urine (32, 33). All these studies were too small to provide useful estimates of the prevalence of Wilson disease and most of them are prone to selection biases that would tend to overestimate the disease prevalence. Thus, the Chinese study (27) involved screening of 153,370 inhabitants in the Hashan County and it is unclear how many declined the invitation. Several of the studies (28-31) included patients at pediatric departments where Wilson disease most likely is more frequent than in the background population. In one study, the selection method was not described (33). In the two studies from Japan (31, 32) including well defined unselected populations of children, the prevalence estimates were either 0:126,810 (31) or 1:11,362 (32), which is hard to conclude from because of large confidence intervals. In summary, these studies do not provide evidence to qualify our discussion.

Population-based genetic studies

We identified 9 studies, where population-based genetic methods were applied to estimate the frequency of Wilson disease-causing alleles. In some studies, the authors only looked for the most frequent known disease-causing mutations and scaled up the estimate to compensate. Other studies looked for all possibly disease-causing mutations.

A Swedish study looked for the H1069Q and T977M mutations that constituted 44% of disease-causing alleles in 24 unrelated Swedish Wilson patients (34). In 2460 blood donors, these alleles together constituted 0.175% (95% CI 0.09% – 0.36%) of *ATP7B* alleles (35). From that a frequency of disease-causing alleles of 0.175%/0.44= 0.39% and a corresponding genetic prevalence of 1:63,175 (CI 1:14,938 – 1:239,012) can be calculated. Only nine mutated alleles were detected in the donor pool and that explains the wide confidence interval.

In a study from the US (36), 2.601 Caucasian newborns were examined for the H1069Q mutation which constitutes 38-45% of Wilson disease-causing mutations in that ethnic group. Seven carriers were found providing a prevalence estimate of 1:50,000 (95% CI 1:18,000-1:700,000). The broad confidence interval shows that even this sample size was too small for firm conclusions.

A study from the UK (37) sequenced the *ATP7B* gene in blood sampled from 1000 newborns (1K sample). They identified 24 alleles with single nucleotide variants predicted by a computer program (GAP4, Staden Sequence Analysis Package) to cause dysfunction of ATP7B. From that observation, they calculated a genetic prevalence of 1:7,026. (95% CI 1:1,800- 1:15,000, our calculation). However, only 12 of the identified mutations were actually known to cause Wilson disease. Recalculating using only known disease causing mutations, the genetic prevalence estimate was 1:28,000 (95% CI: 1:9,100 – 1:104,000, our calculation). It is thus uncertain whether the UK estimate is different from the US estimate of

1:50,000 (36). The UK-paper also included a limited analysis of a larger number of 5,376 samples from newborn individuals (the 5K sample), where they sequenced exons 8, 14, and 18, that cover 51% of all known Wilson causing mutations in the UK. Twenty-three alleles with possible disease-causing mutations were identified, from which we calculated the frequency of disease-causing alleles in that sample to be (23/0.51)/(2*5376)= 0.0042 corresponding to a prevalence of 1:57,000 (CI 1: 35,000 – 1:132,000). This result was again compatible with the US study.

Four studies from South Korea (38-41) analyzed the frequency of the 3, 4, 6, or 7 most common mutations in South Korea in samples from 500(40), 476(39), 14,835(38), and 3,057 individuals(41), respectively. The corresponding genetic prevalence estimates were 1:3,000, 1:30,800, 1:7,500 and 1:27,000 (38-41). These results have been taken to suggest a higher prevalence in South Korea than in the rest of the world, but selection biases may have hampered this conclusion. Thus, in the larger study (38) newborns were recruited only if their parents wanted the test and payed for it, and in two other studies (40, 41), samples were taken from individuals in health promotion centers, where patients tend to accumulate. Unfortunately, there are no clinical data from South Korea to compare with.

In a recent French study (42), ATP7B DNA was sequenced in 697 patients with other diseases than hepatic and neurologic and without a family history of Wilsons disease. Among the 1394 alleles examined, they identified 15 alleles that have classified as pathogenic, 7 "likely pathogenic" alleles and 15 alleles with "variants of uncertain significance". The "likely pathogenic" variants had never been identified in Wilson patients but computer analysis predicted them to be deleterious for protein function. "Variants of uncertain significance" were used when these computer programs produced conflicting predictions in alleles that have never been identified as causing Wilson disease. Combining "pathogenic" and "likely pathogenic", they concluded that 22/1394 (q=0.016, prevalence 1:4014). If only

the 15 "pathogenic" alleles were included, we calculate q=0.011 (0.007-0.018) and prevalence 1:8636 (1:3086-1:20,408) which is still much higher than 1:37,000 we extracted from their clinical study (see above) (20).

As explained above, the genetic prevalence will always be higher than the disease prevalence because of the age at presentation. In the US and British studies, the data on clinical and genetic prevalence did not clearly differ. This was in contrast to the French study, where there seemed to be a significant difference between clinical and genetic prevalence. The same is true for the best studied region, Sardinia to be described below.

A very recent report (43) searched the gnomeAD database of 64,600 non-Finnish European genomes for "pathogenic" and "likely pathogenic" variants in ATP7B, mostly based on computer predictions. Many of the "likely pathogenic" variants had not been observed in WD patients. Gao et al. concluded a prevalence of 1:7,200, close to the French estimate. However, when the most common "pathogenic" or "likely pathogenic" variants from the gnomAD database are compared to the recently published actual distribution of disease causing mutations in 1346 WD patients (44), it is likely that at least a number of these variants are of low penetrance or without relevance for WD at all (Figure 1).

The Sardinian case

The Italian island, Sardinia, is the best studied area employing both clinical and genetic methodologies.

Clinical studies. The first epidemiological study included all identified patients between 1902 and 1983 (45). From the patients alive on certain dates in 1951, 1961, 1971 and 1981, the crude prevalence estimates were 1:58,000, 1:44,000, 1:44,000, and 1:34,000, respectively. In accordance, the number of diagnoses per decade was steadily increasing (45). The prevalence

relative to births was estimated to be 16/ 266,944 = 1:16,684, suggesting that a large number of patients were repeatedly overlooked. In a later study from 2013 (46), 192 patients were included, which corresponds to a prevalence of 1:8,700. The increase of the crude prevalence from 1:58,000 in 1951 to 1:8,700 in 2013 is likely the combined effects of increased diagnostic awareness and improved survival after introduction of penicillamine (45).

Genetic studies. Two population based genetic studies have been performed in Sardinia. In one, 5290 newborns were screened for the typical Sardinian mutation in *ATP7B* (-441/-447 del), and the allele frequency for all Wilson disease-causing mutations in *ATP7B* was calculated to be 1.92 %. Assuming Hardy Weinberg equilibrium, the incidence of Wilson disease should be 0.0192^2 = 1:2,707 of births (14) and even higher in the likely case of Hardy Weinberg violations. A later study using another genetically-based method reached almost the same result (46).

Thus, in Sardinia, like in the French and the gnomAD based studies, there is a clear contrast between the high prevalence predicted in the genetic studies (1:2,700) (14, 43, 46) and the estimates based on clinical diagnoses (1:16,000 and 1:8,700) (45, 46).
One explanation could be that 50-75% of all Wilson patients in Sardinia remained undiagnosed after 30 years of increasing diagnostic awareness, which is hard to believe. Alternatively, the penetrance of Wilson disease-causing mutations in this population is less than 100%.

Discussion

The estimated global prevalence of Wilson disease has not been systematically assessed since 1984 where Scheinberg and Sternlieb suggested 1:30,000 based on the available but limited data at that time. In the present study we reevaluated this estimate including more resent literature and results based on new methodologies.

Since the aim of this study was to estimate the disease prevalence as defined, the clinical epidemiological studies are of key importance. A general tendency in the clinical studies is that the crude prevalence increased with time. Since the diagnostic criteria for Wilson disease have been stable over a long period of time (47, 48), this must be due to factors such as increased diagnostic awareness, earlier diagnosis, shorter diagnostic delay, and the fact that many of the cases overlooked in the past died within a few years of the missed diagnosis and thus exited the calculation.

The larger studies clearly suggest that there are still a number of undiagnosed cases. This is supported by prevalence differences among geographical areas and age groups in the French study (42), the gender differences in the Taiwan study (21), and the increase of prevalence estimates with time in the Hong Kong Study (22). Another factor affecting clinical prevalence is potential masking as other diseases, e.g. as non-alcoholic steatohepatitis or dementia.

When the increasing crude prevalence is considered, interpretation of the four high-quality nationwide studies from Austria, France, Taiwan, and Hong Kong resulted in estimates of the crude disease prevalence ranging from 1:29,000 to 1:40,000 (19-22). Estimates based on the number of diagnosed cases per year relative to the number of births likewise tried to overcome the bias from undiagnosed cases in the past and returned estimates in the same range (Table 1). This is surprisingly close to the original suggestion from 1984.

At the same time some regions may have higher prevalence estimates, such as reported from Croatia (1:28,000), Sardinia (1:16,700-1:34,000), Israel (1:16,000) and Costa Rica (1:19,000). The reason for these differences may be higher frequencies of the disease-causing mutations and/or more first cousin parenthoods. The latter can be quantitatively important; for example, with a carrier frequency of 1:100 an increase of first cousin parenthood frequency from 0 to 5% and 10% will increase the prevalence from 1:40,000 to 1:25,500 and 1:18,700 (Table 3). In a number of Middle East and oriental countries, 25-50 % of all marriages are first cousin relations, and parents of Wilson disease patients are first cousins in up to 54% of cases (49).

Thus, the preliminary conclusion is that in countries with few first cousin marriages, the expected disease prevalence is 1:29,000 -1:40,000, while in the Middle East, Pakistan, India, and other countries where first cousin marriages are common, the disease prevalence will be correspondingly higher (Table 3). Even this conservative estimate suggests a large number of undiagnosed cases worldwide.

Recent technological developments have enabled genetic screening of specific populations, aiming for revised healthcare strategies, based on more precise estimates of the number of undiagnosed Wilson cases. However, in these studies the picture is less clear. While the Swedish and the US studies suggested a genetic prevalence of 1:63,000 to 1:50,000, and the 5K sample of the UK study 1:57,000, the rest of the UK study was somewhat conflicting. In the UK study's 1K sample, detection of known disease-causing mutations suggested a genetic prevalence of 1:28,000 which may not be statistically different from the Swedish and US studies. Both the French and the UK study found very high allele frequencies when they also included computer-predicted pathogenic mutations in their analysis.

In our judgement, to the extent that computer-predicted mutations have not been observed in

Wilson patients, it would be more reasonable to assume that they do not cause Wilson disease. Likewise, the algorithms used for these computer predictions have limitations as the different programs yield conflicting results. Known examples are the common disease-causing Spanish mutation, M645R, predicted to be benign by SIFT, Polyphen, CADD, REVEL MetalR, and Mutation Assessor and the clinically benign mutation, A1140V, predicted by Mutation Taster to be disease-causing with high probability. As genetic prevalence will always be higher than clinical disease prevalence, this consideration further narrows the gap between genetic and high-quality clinical methodologies.

The genetic studies from South Korea could suggest a somewhat higher genetic prevalence in South Korea, but as there are no clinical epidemiological data from that country and since the studies were underpowered or subject to selection biases, no firm conclusions can be made.

The French and Sardinian studies provide challenging data. In both cases, genetic prevalence was approximately 4 times the maximally observed crude prevalence. A recent study (43) of 64,600 genomes from non-Finnish Europeans populations suggested a genetic WD prevalence of 1:7,200 close to the French and British genetic estimates. However, when the frequency of the assumed disease-causing variants in that study was compared to the frequency of disease-causing variants in a large WD population (44) the comparison (Figure 1) strongly suggested that at least some mutations have a low penetrance. Taken together these studies suggest that the penetrance of Wilson disease may actually be lower than 100 % as generally assumed.

The underlying assumptions for the use of genetic prevalence need to be closer examined. One is that any combination of two "disease causing mutations" will lead to Wilson disease. This may not always be the case. It may be that only some but not all combinations lead to Wilson disease. In such case, the genetic prevalence method will overestimate the disease

prevalence. Another possibility is that a combination of mutations that leads to disease in one patient does not have the same effect in another, i.e. the penetrance is not 100%. Since it is common practice to treat asymptomatic siblings identified by genetic screening, it is difficult to judge how often non-penetrance occurs. Case reports describe individuals with two "disease-causing mutations", who never developed symptomatic disease (50, 51), but it is unclear how often that happens. Epigenetic factors have also been shown to modulate the course of Wilson disease and could result in mild or no disease (52). If the penetrance is in fact below 100%, the diagnostic criteria needs revision (47) as will the handling of asymptomatic siblings. From the data presented so far, it is clear that large, population-based clinical studies of high quality and much larger genetic studies are needed to reach sound conclusions on the disease prevalence and genetic penetrance of Wilson disease. As long as the penetrance question is open, population screening of newborns by genetic or biochemical methods should not be encouraged, because it is unclear whether they will develop disease at all.

In conclusion, from the reviewed data, an estimate of disease prevalence of 1:30,000-1:50,000 appears valid for USA, Europe, and Asia and surprisingly close to the 1984 proposals of Scheinberg and Sternlieb (2). There are definitely populations with higher prevalence. One factor is consanguinity that may increase the clinical prevalence considerably and possibly explain a higher disease prevalence in Middle Eastern countries, Pakistan, and India. In others areas, such as Croatia and Sardinia, a higher frequency of disease-causing mutations also contributes. Data suggesting higher genetic prevalence in South Korea are inconclusive and need comparison to clinical presentation. Practically nothing is known about prevalence of Wilson disease in Africa and South America. Available population-based genetic studies in Caucasian populations in some cases show higher

prevalence than clinically based estimates, raising the question whether the penetrance is actually 100% as generally assumed.

References

1. Sternlieb I, Scheinberg IH. Prevention of Wilson's disease in asymptomatic patients. N Engl J Med 1968;278:352-359.

2. Scheinberg IH, Sternlieb I. Wilson's disease. Philadelphia: WB Saunders, 1984.

3. Moller LB, Horn N, Jeppesen TD, Vissing J, Wibrand F, Jennum P, Ott P. Clinical presentation and mutations in Danish patients with Wilson disease. Eur J Hum Genet 2011;19:935-941.

4. Walshe JM, Yealland M. Wilson's disease: the problem of delayed diagnosis. J Neurol Neurosurg Psychiatry 1992;55:692-696.

5. Merle U, Schaefer M, Ferenci P, Stremmel W. Clinical presentation, diagnosis and long-term outcome of Wilson's disease: a cohort study. Gut 2007;56:115-120.

6. Litwin T, Gromadzka G, Czlonkowska A. Gender differences in Wilson's disease. J Neurol Sci 2012;312:31-35.

7. Ferenci P. Wilson’s Disease. Clinical Gastroenterology and Hepatology 2005;3:726-733.

8. Kim EK, Yoo OJ, Song KY, Yoo HW, Choi SY, Cho SW, Hahn SH. Identification of three novel mutations and a high frequency of the Arg778Leu mutation in Korean patients with Wilson disease. Human Mutation 1999;11:275-278.

9. Loudianos G, Dessi V, Lovicu M, Angius A, Figus A, Lilliu F, De Virgiliis S, et al. Molecular characterization of wilson disease in the Sardinian population--evidence of a founder effect. Hum Mutat 1999;14:294-303.

10. Shah AB, Chernov I, Zhang HT, Ross BM, Das K, Lutsenko S, Parano E, et al. Identification and analysis of mutations in the Wilson disease gene (ATP7B): population

frequencies, genotype-phenotype correlation, and functional analyses. Am J Hum Genet 1997;61:317-328.

11. Huster D, Kühne A, Bhattacharjee A, Raines L, Jantsch V, Noe J, Schirrmeister W, et al. Diverse Functional Properties of Wilson Disease ATP7B Variants. Gastroenterology 2012;142:947-956.e945.

12. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998;17:857-872.

13. Dedoussis GV, Genschel J, Sialvera TE, Bochow B, Manolaki N, Manios Y, Tsafantakis E, et al. Wilson disease: high prevalence in a mountainous area of Crete. Ann Hum Genet 2005;69:268-274.

14. Zappu A, Magli O, Lepori MB, Dessi V, Diana S, Incollu S, Kanavakis E, et al. High incidence and allelic homogeneity of Wilson disease in 2 isolated populations: a prerequisite for efficient disease prevention programs. J Pediatr Gastroenterol Nutr 2008;47:334-338.

15. Cocos R, Sendroiu A, Schipor S, Bohiltea LC, Sendroiu I, Raicu F. Genotype-phenotype correlations in a mountain population community with high prevalence of Wilson's disease: genetic and clinical homogeneity. PLoS One 2014;9:e98520.

16. Hofer H, Willheim-Polli C, Knoflach P, Gabriel C, Vogel W, Trauner M, Muller T, et al. Identification of a novel Wilson disease gene mutation frequent in Upper Austria: a genetic and clinical study. J Hum Genet 2012;57:564-567.

17. **Panichareon B, Taweechue K**, Thongnoppakhun W, Aksornworanart M, Pithukpakorn M, Yenchitsomanus PT, Limwongse C, et al. Six novel ATP7B mutations in Thai patients with Wilson disease. European Journal of Medical Genetics 2011;54:103-107.

18. Ferenci P. Regional distribution of mutations of the ATP7B gene in patients with Wilson disease: impact on genetic testing. Hum Genet 2006;120:151-159.

19. Beinhardt S, Leiss W, Stattermayer AF, Graziadei I, Zoller H, Stauber R, Maieron A, et al. Long-term outcomes of patients with Wilson disease in a large Austrian cohort. Clin Gastroenterol Hepatol 2014;12:683-689.

20. Poujois A, Woimant F, Samson S, Chaine P, Girardot-Tinant N, Tuppin P. Characteristics and prevalence of Wilson's disease: A 2013 observational population-based study in France. Clin Res Hepatol Gastroenterol 2017.

21. Tai CS, Wu JF, Chen HL, Hsu HY, Chang MH, Ni YH. Modality of treatment and potential outcome of Wilson disease in Taiwan: A population-based longitudinal study. J Formos Med Assoc 2018;117:421-426.

22. Cheung KS, Seto WK, Fung J, Mak LY, Lai CL, Yuen MF. Epidemiology and natural history of Wilson's disease in the Chinese: A territory-based study in Hong Kong between 2000 and 2016. World J Gastroenterol 2017;23:7716-7726.

23. Ferenci P, Stremmel W, Czlonkowska A, Szalay F, Viveiros A, Stattermayer AF, Bruha R, et al. Age,sex, but not ATP7B genotype effectively influences the clinical phenotype of Wilson disease. Hepatology 2018.

24. Esezobor CI, Banjoko N, Rotimi-Samuel A, Lesi FE. Wilson disease in a Nigerian child: a case report. J Med Case Rep 2012;6:200.

25. Longe AC, Glew RH, Omene JA. Wilson's disease. Report of a case in a Nigerian. Arch Neurol 1982;39:129-130.

26. Ferenci P, Eurowilson Study Group. Prospective Study of the incidence of Wilson Disease in Europe - The Eurowilson Study [Abstract]. J Hepatol 2008;48:1.

27. Cheng N, Wang K, Hu W, Sun D, Wang X, Hu J, Yang R, et al. Wilson disease in the South chinese han population. Can J Neurol Sci 2014;41:363-367.

28. Hahn SH, Lee SY, Jang YJ, Kim SN, Shin HC, Park SY, Han HS, et al. Pilot study of mass screening for Wilson's disease in Korea. Mol Genet Metab 2002;76:133-136.

29. Kroll CA, Ferber MJ, Dawson BD, Jacobson RM, Mensink KA, Lorey F, Sherwin J, et al. Retrospective determination of ceruloplasmin in newborn screening blood spots of patients with Wilson disease. Mol Genet Metab 2006;89:134-138.

30. Ohura T, Abukawa D, Shiraishi H, Yamaguchi A, Arashima S, Hiyamuta S, Tada K, et al. Pilot study of screening for Wilson disease using dried blood spots obtained from children seen at outpatient clinics. J Inherit Metab Dis 1999;22:74-80.

31. Yamaguchi Y, Aoki T, Arashima S, Ooura T, Takada G, Kitagawa T, Shigematsu Y, et al. Mass screening for Wilson's disease: results and recommendations. Pediatr Int 1999;41:405-408.

32. Nakayama K, Kubota M, Katoh Y, Sawada Y, Saito A, Nishimura K, Katsura E, et al. Early and presymptomatic detection of Wilson's disease at the mandatory 3-year-old medical health care examination in Hokkaido Prefecture with the use of a novel automated urinary ceruloplasmin assay. Mol Genet Metab 2008;94:363-367.

33. Owada M, Suzuki K, Fukushi M, Yamauchi K, Kitagawa T. Mass screening for Wilson's disease by measuring urinary holoceruloplasmin. J Pediatr 2002;140:614-616.

34. Waldenstrom E, Lagerkvist A, Dahlman T, Westermark K, Landegren U. Efficient detection of mutations in Wilson disease by manifold sequencing. Genomics 1996;37:303-309.

35. Olsson C, Waldenstrom E, Westermark K, Landegre U, Syvanen AC. Determination of the frequencies of ten allelic variants of the Wilson disease gene (ATP7B), in pooled DNA samples. Eur J Hum Genet 2000;8:933-938.

36. Olivarez L, Caggana M, Pass KA, Ferguson P, Brewer GJ. Estimate of the frequency of Wilson's disease in the US Caucasian population: a mutation analysis approach. Ann Hum Genet 2001;65:459-463.

37. **Coffey AJ, Durkie M**, Hague S, McLay K, Emmerson J, Lo C, Klaffke S, et al.

A genetic study of Wilson's disease in the United Kingdom. Brain 2013;136:1476-1487.

38. Jang JH, Lee T, Bang S, Kim YE, Cho EH. Carrier frequency of Wilson's disease in the Korean population: a DNA-based approach. J Hum Genet 2017.

39. **Kim GH, Yang JY,** Park JY, Lee JJ, Kim JH, Yoo HW. Estimation of Wilson's disease incidence and carrier frequency in the Korean population by screening ATP7B major mutations in newborn filter papers using the SYBR green intercalator method based on the amplification refractory mutation system. Genet Test 2008;12:395-399.

40. Park HD, Ki CS, Lee SY, Kim JW. Carrier frequency of the R778L, A874V, and N1270S mutations in the ATP7B gene in a Korean population. Clin Genet 2009;75:405-407.

41. Song MJ, Lee ST, Lee MK, Ji Y, Kim JW, Ki CS. Estimation of carrier frequencies of six autosomal-recessive Mendelian disorders in the Korean population. J Hum Genet 2012;57:139-144.

42. Collet C, Laplanche JL, Page J, Morel H, Woimant F, Poujois A. High genetic carrier frequency of Wilson's disease in France: discrepancies with clinical prevalence. BMC Med Genet 2018;19:143.

43. **Gao J, Brackley S**, Mann JP. The global prevalence of Wilson disease from next-generation sequencing data. Genet Med 2019;21:8.

44. Ferenci P, Stremmel W, Czlonkowska A, Szalay F, Viveiros A, Stattermayer AF, Bruha R, et al. Age and Sex but Not ATP7B Genotype Effectively Influence the Clinical Phenotype of Wilson Disease. Hepatology 2019;69:1464-1476.

45. Giagheddu A, Demelia L, Puggioni G, Nurchi AM, Contu L, Pirari G, Deplano A, et al. Epidemiologic study of hepatolenticular degeneration (Wilson's disease) in Sardinia (1902-1983). Acta Neurol Scand 1985;72:43-55.

46. Gialluisi A, Incollu S, Pippucci T, Lepori MB, Zappu A, Loudianos G, Romeo

G. The homozygosity index (hi) approach reveals high allele frequency for wilson disease in the sardinian population. European Journal of Human Genetics 2013;21:1308-1311.

47. Ferenci P, Caca K, Loudianos G, Mieli-Vergani G, Tanner S, Sternlieb I, Schilsky M, et al. Diagnosis and phenotypic classification of Wilson disease. Liver Int 2003;23:139-142.

48. Scheinberg IH, Sternlieb I. Wilson's Disease. Annu Rev Med 1965;16:119-134.

49. Taly AB, Meenakshi-Sundaram S, Sinha S, Swamy HS, Arunodaya GR. Wilson disease: description of 282 patients evaluated over 3 decades. Medicine (Baltimore) 2007;86:112-121.

50. Hefter H, Weiss P, Wesch H, Stremmel W, Feist D, Freund HJ. Late diagnosis of Wilson's disease in a case without onset of symptoms. Acta Neurol Scand 1995;91:302-305.

51. Czlonkowska A, Rodo M, Gromadzka G. Late onset Wilson's disease: therapeutic implications. Mov Disord 2008;23:896-898.

52. Kieffer DA, Medici V. Wilson disease: At the crossroads between genetics and epigenetics-A review of the evidence. Liver Res 2017;1:121-130.

53. Tschumi A, Colombo JP, Moser H. [Wilson's disease in Switzerland. Clinical, genetic and biochemical studies]. Schweiz Med Wochenschr 1973;103:140-145 concl.

54. Bachmann H, Lossner J, Gruss B, Ruchholtz U. [The epidemiology of Wilson's disease in the German Democratic Republic and current problems from the viewpoint of population genetics]. Psychiatr Neurol Med Psychol (Leipz) 1979;31:393-400.

55. Przuntek H, Hoffmann E. [Epidemiologic study of Wilson's disease in West Germany]. Nervenarzt 1987;58:150-157.

56. O'Brien M, Reilly M, Sweeney B, Walsh C, Hutchinson M. Epidemiology of Wilson's disease in Ireland. Mov Disord 2014;29:1567-1568.

57. Almdal TP, Sorensen TI. Incidence of parenchymal liver diseases in Denmark, 1981 to 1985: analysis of hospitalization registry data. The Danish Association for the Study of the Liver. Hepatology 1991;13:650-655.

58. Moller LB, Ott P, Lund C, Horn N. Homozygosity for a gross partial gene deletion of the C-terminal end of ATP7B in a Wilson patient with hepatic and no neurological manifestations. Am J Med Genet A 2005;138:340-343.

59. Bruha R, Marecek Z, Pospisilova L, Nevsimalova S, Vitek L, Martasek P, Nevoral J, et al. Long-term follow-up of Wilson disease: natural history, treatment, mutations analysis and phenotypic correlation. Liver Int 2011;31:83-91.

60. Naorniakowska M, Dadalski M, Kaminska D, Janczyk W, Lebensztejn D, Fyderek K, Wysocki J, et al. Clinical presentations of Wilson disease among Polish children. Dev Period Med 2016;20:216-221.

Author names in bold designate shared first co-authorship.

Figure legend

Figure 1. Comparison of predicted and observed disease-causing variants of ATP7B.

The X-axis displays the frequency of the most common predicted “pathogenic” or “likely pathogenic” variants of ATP7B in the gnomAD database (https://gnomad.broadinstitute.org/) according to Gao et al. (43). The Y-axis displays the frequency of disease-causing mutations in 1346 European Wilson patients (44). On the assumptions that (A) the non-Finnish samples in the gnomAD database is a relevant background population for patient database, (B) Hardy Weinberg and (C) 100% penetrance, all data points should be on a straight line. If one assumes that the penetrance of the R969Q variant is 100% data should be on the line depicted

on the graph. A number of mutations are actually close to that line and could represent mutations with (close to) 100% penetrance. Other data points are on the graph but clearly below that line. They could represent mutations with lower than 100% penetrance. Still a number of variants - those placed at the X-axis - were not present in the patient database at all. They may not cause Wilson disease.

Table 1

Prevalence estimates based on the number of presenting cases relative to number of births.

Country and period	**Cases diagnosed**	**No. of births**	**Prevalence estimate**	**Ref**
Switzerland 1946-55	19		1:44,800	(53)
Eastern Germany 1949-1977	123	4.2 mio/28 years	1:34,400	(54)
Western Germany 1962			1:86,000	(55)
Sardinia 1971-1981	16	266,944	1:16,700	(45)
Ireland 1980-1989	12	637	1:53,000	(56)
Denmark 1981-1985	4	55,000/year	1:68,750	(57)
Denmark 1990-2009	28	70,000/year	1:49,500	(58)
Czech Republic 1995-2008	76	95,000/year	1:16,500	(59)
Poland 1996-2016	156	385,000/year	1:49,000	(60)

Table 2

Screening for Wilson Disease by measurement of ceruloplasmin in blood or urine or detection of the Kayser-Fleisher (KF)-ring.

Method/year	Country	No	WD	Prevalence	Ref
Blood Ceruloplasmin					
1999	Japan	126,810	0	0	(31)
1999	Japan	24,165	3	1:8,084	(31)
1999	Japan	2,789	2	1:1,400	(30)
2002	South Korea	3,667	1	1:3,667	(28)
2006	USA	1,398	0	0	(29)
U-Holo-Ceruloplasmin					
2002	Japan	48,819	2	1:24,400	(33)
U-Ceruloplasmin					
2008	Japan	11,362	1	1:11,362	(32)
KF-Rings					
2014	China	153,370	9	1:17.000	(27)

Table 3

Predicted prevalence of Wilson disease as a function of the frequency of disease-causing alleles, q, and percentage of first cousin marriages in a theoretical population. Full penetrance and normal life expectancy are assumed and other sources of consanguinity are neglected.

q	Carrier frequency	Percentage of first cousin marriages				
		0	**0.05**	**0.1**	**0.25**	**0.5**
1:300	1:151	1:90,000	1:47,761	1:32,506	1:16,599	1:9,143
1:200	1:101	1:40,000	1:25,447	1:18,659	1:10,364	1:5,953
1:100	1:51	1:10,000	1:7,940	1:6,584	1:4,354	1:2,783
1:50	1:26	1:2,500	1:2,266	1:2,073	1:1,649	1:1,231