

EDITORIAL

Open Access



The Prospective Lynch Syndrome Database: background, design, main results and complete MySQL code

Pål Møller^{1,2*}

Abstract

A brief description of why and for which purposes the Prospective Lynch Syndrome Database was established, the principles and design, and the main classes of results are given. Data input is assumption-free input enabling validation of paradigms used to explain the results. The design is considering cancer/age as discrete events to occur or not in a time dimension in a closed room compliant with population genetic paradigms and last centuries developing paradigms of interpreting discrete events reflecting conditional and/or co-occurring stochastic probabilities. Which may be in contrast to the paradigm that any observed event has a cause. The results may indicate that some current paradigms on carcinogenesis should be reconsidered. The complete analytic code in MySQL© syntax together with a flowchart illustrating how the different pieces of codes interrelate are included as supplementary files, enabling third parties to use or modify the code to examine prospectively observed events in their own activities when referring to this report as the source.

The reason for making the Prospective Lynch Syndrome Database (PLSD) was that ten years ago it was obvious that the existing paradigms did not completely explain what was being observed: Variants in each of the MMR genes causing Lynch syndrome (LS) had different but uncertain penetrance (cumulative incidence of cancer) and expressivities (organs in which cancer occurred). Because no one has an “average genetic variant” or an ‘average gender’, the “average” penetrance and expressivity were valid for no one. It was known that colonoscopy did not prevent colon cancer as had been hoped (which is why the advocated interval between colonoscopies was reduced in updated clinical guidelines), but the true incidence of colon cancer in those at risk who had regular colonoscopy was not known. The PLSD was designed

to describe penetrance and expressivities of cancer by age, gene and gender, survival when cancer occurred and survival in healthy carriers of pathogenic MMR variants when interventions including colonoscopy interfered with the natural history associated with LS.

Empirical evidence is needed from prospective studies to obtain more accurate information on prospective incidences of phenotypes. Three examples why retrospective information may not predict future risks for inherited phenotypes may be: (1) In a stable population, 33% of woman will have a sister when complete selection ascertainment, while the probability for an expecting woman that her baby will be a girl is about 50% irrespective of any previous girls she may have delivered. (2) With multi-genic inheritance, the next generation will have a reduced phenotypic incidence. (3) Retrospective ascertainment biases may cause an artificial increase in phenotypic incidence (anticipation).

When validating paradigms, no information entered into the database should be restricted, ascertained, conditioned or based upon the paradigms that are to be

*Correspondence: moller.pal@gmail.com

¹ Department of Tumor Biology, Institute of Cancer Research, The Norwegian Radium Hospital, 0379 Oslo, Norway
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

validated (you cannot have your ascertainment parameter as your study parameter). The design of the study should not be restricted to the examination of one specific paradigm or claim – if restricting your design to examine one hypothesis, the results will be nothing but a validation of that specific hypothesis. In the context of LS, if two or more carcinogenetic pathways may be operating simultaneously, proving an adenoma to be a precursor of cancer does not exclude other carcinogenetic pathways to colon cancer. The idea of the PLSD was to measure by how much colonoscopy reduced colorectal cancer incidence – it came as a surprise that colorectal cancer was apparently not reduced as a result of colonoscopy [1]. The design of the PLSD, however, did not include any initial paradigm and the design allowed unexpected results: The PLSD reports on the observed empirical evidence based on assumption-free data. This does not, however, exclude ascertainment biases which any study will have and in relation to which results must be appropriately interpreted [2].

Most studies of carcinogenesis and cancer clinical trials explore or test one or a few hypotheses, the ‘gold standard in oncological research’ of a randomized trial being the classical example. In doing so, you will find nothing besides what you are looking for: you may describe one tree in detail, but you may not see the forest.

PLSD was designed as follows: The primary object (never to be duplicated and never to be split into two) is the carrier, and to the carriers are attached the attributes gender and genetic variant which are inborn and will never change. The events recorded are any cancer/age when diagnosed, age at inclusion for follow-up, age at last observation, and age at death. The analytic model needs an adequate number of carriers included in all age groups, and especially so in the youngest age groups when calculating cumulative incidences. Power analyses indicated that about 2,000 carriers should be included for a first analysis that did not stratify on gender [3]. As, at the time, no single centre or even country had enough carriers to provide the required granularity in their data, in 2012 members of the Mallorca Group (now EHTG www.ehtg.org) agreed to compile data on their prospectively observed carriers. The increased number of cases later included allowed more detailed analyses.

To handle such numbers of cases, a computerized database is required: The basic principles of the PLSD relational database and how it was de-normalized for the current task have been previously described [4, 5]. Also, the major confounders to be considered when interpreting the results in oncological research like this have been previously discussed [2]. The basic PLSD administrative description and data call is available online <https://www.ehtg.org/plsd.php>. While the PLSD

reports present cumulative risk for cancers by gene and gender from 25 years of age onwards, cumulative risks starting from any current age of 25 years or more may be calculated interactively at www.plsd.eu, based on the last PLSD report. This may be of interest to carriers of different ages.

The PLSD reports so far indicate that each of the four LS genes when deranged are associated with different inherited cancer syndromes. All cancers included in the LS spectrum (except for brain tumours and osteosarcomas) are derived from the embryonic endothelial tissues, but occur with different incidences in the different affected organs according to the respective germline pathogenic variant.

- The MSH2 syndrome is a truly multi-organ dominantly inherited cancer syndrome frequently having extra-intestinal cancers in the urinary tract, prostate and brain and in women frequent endometrial and ovarian cancer. Colonoscopy may be associated with over-diagnosis of colorectal cancer.
- The MLH1 syndrome has high incidences of gynecological, colorectal and upper intestinal cancers with a higher incidence of colorectal cancer in males than in females. Colonoscopy may be associated with over-diagnosis of colorectal cancer.
- The MSH6 syndrome is by and large a dominantly inherited sex-limited cancer syndrome having a high incidence for endometrial or ovarian cancer and a slightly increased and gender-equal incidence of colorectal cancer. It is under-reported because males often appear as skipped generations in families. There is no evidence that colonoscopy reduces incidence of colorectal cancer.
- The PMS2 syndrome has lower but increased incidence of endometrial and colorectal cancer but seemingly no validated increased risk for cancer in other organs. It is under-reported because low penetrance makes it difficult to delineate from normal variation. In contrast to the others, colonoscopy may reduce the incidence of colorectal cancer before 50 years of age.

Ten years survival following early diagnosis and treatment of colorectal or endometrial cancers in carriers of pathogenic variants of *MSH2* or *MLH1* is high. For carriers of pathogenic variants of all four genes who are subjected to follow-up as advocated by international guidelines, death from extra-colorectal cancers now seems more frequent than from colorectal cancers, especially so for carriers of pathogenic variants of *MSH2* and *MSH6*. *MLH1* carriers may die from later pancreatic or biliary tract cancers. Risk from dying

from ovarian cancer diagnosed before 40 years of age for carriers of any genetic variant was calculated to 0%, concluding that prophylactic oophorectomy before that age may not be indicated [6].

Making the PLSD MySQL algorithms available to all should facilitate others in validating the published PLSD results, which is a general requirement for trusting new information that is used for health care. For the reason the algorithms are here made available for all in the [supplementary files](#): The complete MySQL codes for all tables, views and functions; a flow-chart on how the different parts interact; installation notes; and a few cases to populate the tables as a demonstration for a start. It runs on the free version of MySQL8®. Tables may be populated by importing the data from any source having an ODBC driver; any of the final outputs will be displayed within a second or two upon request and results may be exported for further processing to any other application having an ODBC driver. The MySQL outputs are in three classes: (1) Number of events, number follow-up years and the fraction as annual incidence rates in each age cohort for any filtering selection as specified; (2) output tailored for K-M survival analyses when certain events are under consideration; and (3) some counts of cases included by carrier status and geographical region lived in. Additional tables including more attributes of the carriers may be added to the MySQL database, and additional views/functions may be added to answer additional questions and to analyze more substrata. Further calculations of cumulative incidence should be based on Poisson distributions as previously published [1].

The design of the PLSD database (cancers considered as discrete events by age) is compliant with the requirements included in population genetic theories, including closed room statistics like hypergeometric distributions, Boolean parameters and Bayesian probability calculations, etc. when appropriate. The design is also compliant with the new paradigms on stochastic probabilities as causes for events, which may be considered in contrast to Newton's paradigm of any event having a specific cause, and this may be of interest for understanding carcinogenetic mechanisms.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13053-022-00243-z>.

Additional file 1.

Acknowledgements

PM designed the PLSD and wrote the MySQL code and has IPR to the PLSD code included. The PLSD is a group effort based on support from the carriers described, contributors of data, and the support of the EHTG board members.

A special thank you to the writing team for the PLSD reports: Mev Dominguez-Valentin (curator), Toni Säppälä and Julian Sampson.

Waiver of responsibility

PM will have no responsibility for results using the code here published, or from interpretation of such results.

Authors' contributions

Not applicable. The author(s) read and approved the final manuscript.

Funding

No separate funding.

Availability of data and materials

All included in supplementary files.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

No competing interests.

Author details

¹Department of Tumor Biology, Institute of Cancer Research, The Norwegian Radium Hospital, 0379 Oslo, Norway. ²The European Hereditary Tumour Group, Edinburgh, Scotland, UK.

Received: 24 October 2022 Accepted: 9 November 2022

Published online: 21 November 2022

References

- Møller P, Seppälä T, Dowty JG, Haupt S, Dominguez-Valentin M, Sunde L, Bernstein I, Engel C, Aretz S, Nielsen M, Capella G, Evans DG, Burn J, Holinski-Feder E, Bertario L, Bonanni B, Lindblom A, Levi Z, Macrae F, Winship I, Plazzer JP, Sijmons R, Laghi L, Valle AD, Heinimann K, Half E, Lopez-Koestner F, Alvarez-Valenzuela K, Scott RJ, Katz L, Laish I, Vainer E, Vaccaro CA, Carraro DM, Gluck N, Abu-Freha N, Stakelum A, Kennelly R, Winter D, Rossi BM, Greenblatt M, Bohorquez M, Sheth H, Tibiletti MG, Lino-Silva LS, Horisberger K, Portenkirchner C, Nascimento I, Rossi NT, da Silva LA, Thomas H, Zaránd A, Mecklin JP, Pylvänäinen K, Renkonen-Sinisalo L, Lepistö A, Peltomäki P, Therkildsen C, Lindberg LJ, Thorlacius-Ussing O, von Knebel Doeberitz M, Loeffler M, Rahner N, Steinke-Lange V, Schmiegel W, Vangala D, Perne C, Hüneburg R, de Vargas AF, Latchford A, Gerdes AM, Backman AS, Guillén-Ponce C, Snyder C, Lautrup CK, Amor D, Palmero E, Stoffel E, Duijkers F, Hall MJ, Hampel H, Williams H, Okkels H, Lubiński J, Reece J, Ngeow J, Guillem JG, Arnold J, Wadt K, Monahan K, Senter L, Rasmussen LJ, van Hest LP, Ricciardiello L, Kohonen-Corish MRJ, Ligtenberg MJL, Southey M, Aronson M, Zahary MN, Samadder NJ, Poplawski N, Hoogerbrugge N, Morrison PJ, James P, Lee G, Chen-Shtoyerman R, Ankathil R, Pai R, Ward R, Parry S, Dębniak T, John T, van Overeem Hansen T, Caldés T, Yamaguchi T, Barca-Tierno V, Garre P, Cavestro GM, Weitz J, Redler S, Büttner R, Heuveline V, Hopper JL, Win AK, Lindor N, Gallinger S, Le Marchand L, Newcomb PA, Figueroa J, Buchanan DD, Thibodeau SN, Ten Broeke SW, Hovig E, Nakken S, Pineda M, Dueñas N, Brunet J, Green K, Lalloo F, Newton K, Crosbie EJ, Mints M, Tjandra D, Neffa F, Esperon P, Kariv R, Rosner G, Pavicic WH, Kalfayan P, Torrezan GT, Bassaneze T, Martin C, Moslein G, Ahadova A, Kloor M, Sampson JR, Jenkins MA, European Hereditary Tumour Group (EHTG) and the International Mismatch Repair Consortium (IMRC). Colorectal cancer incidences in Lynch syndrome: a comparison of results from the prospective lynch syndrome database and the international mismatch repair consortium. *Hered Cancer Clin Pract.* 2022;20(1):36. <https://doi.org/10.1186/s13053-022-00241-1> PMID:36182917; PMCID: PMC9526951.
- Møller P. The Prospective Lynch Syndrome Database reports enable evidence-based personal precision health care. *Hered Cancer Clin Pract.*

- 2020;14:6. <https://doi.org/10.1186/s13053-020-0138-0> PMID: 32190163; PMCID: PMC7073013.
3. Møller P, Seppälä T, Bernstein I, Holinski-Feder E, Sala P, Evans DG, Lindblom A, Macrae F, Blanco I, Sijmons R, Jeffries J, Vasen H, Burn J, Nakken S, Hovig E, Rødland EA, Tharmaratnam K, de VTotN Cappel, Hill WH, Wijnen J, Green J, Lalloo K, Sunde F, Mints L, Bertario M, Pineda L, Navarro M, Morak M, Renkonen-Sinisalo M, Frayling L, Plazzer IM, Pylvanainen JP, Sampson K, Capella JR, Mecklin G, Möslein JP, Mallorca G, Group. <http://mallorca-group.eu>. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. *Gut*. 2017 Mar;66(3):464–72. doi:<https://doi.org/10.1136/gutjnl-2015-309675>. Epub 2015 Dec 9. PMID: 26657901; PMCID: PMC5534760.
 4. Møller P, Nakken S, Hovig E. Databases: Intentions, Capabilities, and Limitations. In: Valle L, Gruber S, Capellá G, editors. *Hereditary Colorectal Cancer*. Cham: Springer; 2018. https://doi.org/10.1007/978-3-319-74259-5_26.
 5. Møller P, Nakken S, Hovig E. The Prospective Lynch Syndrome Database. In: Valle L, Gruber S, Capellá G, editors. *Hereditary Colorectal Cancer*. Cham: Springer; 2018. https://doi.org/10.1007/978-3-319-74259-5_28.
 6. Dominguez-Valentin M, Crosbie EJ, Engel C, Aretz S, Macrae F, Winship I, Capella G, Thomas H, Nakken S, Hovig E, Nielsen M, Sijmons RH, Bertario L, Bonanni B, Tibiletti MG, Cavestro GM, Mints M, Gluck N, Katz L, Heinimann K, Vaccaro CA, Green K, Lalloo F, Hill J, Schmiegel W, Vangala D, Perne C, Strauß HG, Tecklenburg J, Holinski-Feder E, Steinke-Lange V, Mecklin JP, Plazzer JP, Pineda M, Navarro M, Vidal JB, Kariv R, Rosner G, Piñero TA, Gonzalez ML, Kalfayan P, Ryan N, Ten Broeke SW, Jenkins MA, Sunde L, Bernstein I, Burn J, Greenblatt M, de Vos Tot Nederveen Cappel WH, Della Valle A, Lopez-Koestner F, Alvarez K, Büttner R, Görgens H, Morak M, Holzapfel S, Hüneburg R, von Knebel Doeberitz M, Loeffler M, Rahner N, Weitz J, Pylvänäinen K, Renkonen-Sinisalo L, Lepistö A, Auranen A, Hopper JL, Win AK, Haile RW, Lindor NM, Gallinger S, Le Marchand L, Newcomb PA, Figueiredo JC, Thibodeau SN, Therkildsen C, Okkels H, Ketabi Z, Denton OG, Rødland EA, Vasen H, Neffa F, Esperon P, Tjandra D, Möslein G, Sampson JR, Evans DG, Seppälä TT, Møller P. Risk-reducing hysterectomy and bilateral salpingo-oophorectomy in female heterozygotes of pathogenic mismatch repair variants: a Prospective Lynch Syndrome Database report. *Genet Med*. 2021;23(4):705–12. <https://doi.org/10.1038/s41436-020-01029-1> Epub 2020 Dec 1. PMID: 33257847; PMCID: PMC8026395.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

