Translational Artificial Intelligence (AI): The Need to Translate from Basic Science to Clinical Value

William Hersh, MD
Professor
Department of Medical Informatics & Clinical Epidemiology
School of Medicine
Oregon Health & Science University
Portland, OR, USA
https://www.ohsu.edu/informatics
Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: https://informaticsprofessor.blogspot.com/
Twitter: @williamhersh

References

Alexander, L., 2016. Do Google's "unprofessional hair" results show it is racist? the Guardian.

Attia, Z.I., Friedman, P.A., Noseworthy, P.A., Lopez-Jimenez, F., Ladewig, D.J., Satam, G., Pellikka, P.A., Munger, T.M., Asirvatham, S.J., Scott, C.G., Carter, R.E., Kapa, S., 2019. Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs. Circ Arrhythm Electrophysiol 12, e007284. https://doi.org/10.1161/CIRCEP.119.007284

Balwani, M., Sardh, E., Ventura, P., Peiró, P.A., Rees, D.C., Stölzel, U., Bissell, D.M., Bonkovsky, H.L., Windyga, J., Anderson, K.E., Parker, C., Silver, S.M., Keel, S.B., Wang, J.-D., Stein, P.E., Harper, P., Vassiliou, D., Wang, B., Phillips, J., Ivanova, A., Langendonk, J.G., Kauppinen, R., Minder, E., Horie, Y., Penz, C., Chen, J., Liu, S., Ko, J.J., Sweetser, M.T., Garg, P., Vaishnaw, A., Kim, J.B., Simon, A.R., Gouya, L., ENVISION Investigators, 2020. Phase 3 Trial of RNAi Therapeutic Givosiran for Acute Intermittent Porphyria. N Engl J Med 382, 2289–2301. https://doi.org/10.1056/NEJMoa1913147

Bejnordi, B.E., Zuidhof, G., Balkenhol, M., Hermsen, M., Bult, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J., 2017. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. J Med Imaging (Bellingham) 4, 044504. https://doi.org/10.1117/1.JMI.4.4.044504

Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine 25, 1301–1309. https://doi.org/10.1038/s41591-019-0508-1

Carlini, N., 2020. Privacy Considerations in Large Language Models. Google AI Blog. URL http://ai.googleblog.com/2020/12/privacy-considerations-in-large.html (accessed 3.18.21).

Chamberlin, S.R., Bedrick, S.D., Cohen, A.M., Wang, Y., Wen, A., Liu, S., Liu, H., Hersh, W.R., 2020. Evaluation of patient-level retrieval from electronic health record data for a cohort discovery task. JAMIA Open 3, 395–404. https://doi.org/10.1093/jamiaopen/ooaa026

Chi, E.A., Chi, G., Tsui, C.T., Jiang, Y., Jarr, K., Kulkarni, C.V., Zhang, M., Long, J., Ng, A.Y., Rajpurkar, P., Sinha, S.R., 2021. Development and Validation of an Artificial Intelligence

System to Optimize Clinician Review of Patient Records. JAMA Netw Open 4, e2117391. https://doi.org/10.1001/jamanetworkopen.2021.17391

Cohen, A.M., Chamberlin, S., Deloughery, T., Nguyen, M., Bedrick, S., Meninger, S., Ko, J.J., Amin, J.J., Wei, A.J., Hersh, W., 2020. Detecting rare diseases in electronic health records using machine learning and knowledge engineering: Case study of acute hepatic porphyria. PLoS ONE 15, e0235574. https://doi.org/10.1371/journal.pone.0235574

DeCamp, M., Lindvall, C., 2020. Latent bias and the implementation of artificial intelligence in medicine. J Am Med Inform Assoc 27, 2020–2023. https://doi.org/10.1093/jamia/ocaa094

Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., Socher, R., 2021. Deep learning-enabled medical computer vision. npj Digital Medicine 4, 1–9. https://doi.org/10.1038/s41746-020-00376-2

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118. https://doi.org/10.1038/nature21056

Flanagin, A., Bibbins-Domingo, K., Berkwits, M., Christiansen, S.L., 2023. Nonhuman "Authors" and Implications for the Integrity of Scientific Publication and Medical Knowledge. JAMA. https://doi.org/10.1001/jama.2023.1344

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.L., Webster, D.R., 2016. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 316, 2402–2410. https://doi.org/10.1001/jama.2016.17216

Haendel, M., Vasilevsky, N., Unni, D., Bologa, C., Harris, N., Rehm, H., Hamosh, A., Baynam, G., Groza, T., McMurry, J., Dawkins, H., Rath, A., Thaxon, C., Bocci, G., Joachimiak, M.P., Köhler, S., Robinson, P.N., Mungall, C., Oprea, T.I., 2020. How many rare diseases are there? Nat Rev Drug Discov 19, 77–78. https://doi.org/10.1038/d41573-019-00180-y

Haenssle, H.A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A.B.H., Thomas, L., Enk, A., Uhlmann, L., Reader study level-I and level-II Groups, Alt, C., Arenbergerova, M., Bakos, R., Baltzer, A., Bertlich, I., Blum, Andreas, Bokor-Billmann, T., Bowling, J., Braghiroli, N., Braun, R., Buder-Bakhaya, K., Buhl, Timo, Cabo, H., Cabrijan, L., Cevic, N., Classen, A., Deltgen, D., Fink, Christine, Georgieva, I., Hakim-Meibodi, L.-E., Hanner, S., Hartmann, F., Hartmann, J., Haus, G., Hoxha, E., Karls, R., Koga, H., Kreusch, J., Lallas, A., Majenka, P., Marghoob, A., Massone, C., Mekokishvili, L., Mestel, D., Meyer, V., Neuberger, A., Nielsen, K., Oliviero, M., Pampena, R., Paoli, J., Pawlik, E., Rao, B., Rendon, A., Russo, T., Sadek, A., Samhaber, K., Schneiderbauer, Roland, Schweizer, A., Toberer, Ferdinand, Trennheuser, L., Vlahova, L., Wald, A., Winkler, J., Wölbing, P., Zalaudek, I., 2018. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 29, 1836–1842. https://doi.org/10.1093/annonc/mdy166

Hersh, W., 2020. Information Retrieval: A Biomedical and Health Perspective, 4th ed, Health Informatics. Springer International Publishing.

Hersh, W.R., 2022. Health Informatics: Practical Guide, 8th Edition. Lulu.com.

Hersh, W.R., Cohen, A.M., Nguyen, M.M., Bensching, K.L., Deloughery, T.G., 2022. Clinical study applying machine learning to detect a rare disease: results and lessons learned. JAMIA Open 5, ooac053. https://doi.org/10.1093/jamiaopen/ooac053

Hersh, W.R., Gorman, P.N., Biagioli, F.E., Mohan, V., Gold, J.A., Mejicano, G.C., 2014. Beyond information retrieval and electronic health record use: competencies in clinical informatics for medical education. Adv Med Educ Pract 5, 205–212. https://doi.org/10.2147/AMEP.S63903

Hersh, W.R., Weiner, M.G., Embi, P.J., Logan, J.R., Payne, P.R.O., Bernstam, E.V., Lehmann, H.P., Hripcsak, G., Hartzog, T.H., Cimino, J.J., Saltz, J.H., 2013. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care 51, S30-37. https://doi.org/10.1097/MLR.0b013e31829b1dbd

Hoyt, R., 2022a. Maximizing Orange for Data Science Education — Part 1. MLearning.ai. URL https://medium.com/mlearning-ai/maximizing-orange-for-data-science-education-part-1-1be42362634a (accessed 2.1.23).

Hoyt, R., 2022b. Maximizing Orange for Data Science Education — Part 2. MLearning.ai. URL https://medium.com/mlearning-ai/maximizing-orange-for-data-science-education-part-2-7d036c9d8038 (accessed 2.1.23).

Ibrahim, H., Liu, X., Rivera, S.C., Moher, D., Chan, A.-W., Sydes, M.R., Calvert, M.J., Denniston, A.K., 2021. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. Trials 22, 11. https://doi.org/10.1186/s13063-020-04951-6

James, C.A., Wachter, R.M., Woolliscroft, J.O., 2022. Preparing Clinicians for a Clinical World Influenced by Artificial Intelligence. JAMA 327, 1333–1334. https://doi.org/10.1001/jama.2022.3580

Jha, S., Topol, E.J., 2018. Information and Artificial Intelligence. J Am Coll Radiol 15, 509–511. https://doi.org/10.1016/j.jacr.2017.12.025

Jha, S., Topol, E.J., 2016. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. JAMA 316, 2353–2354. https://doi.org/10.1001/jama.2016.17438

Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., Tseng, V., 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health 2, e0000198. https://doi.org/10.1371/journal.pdig.0000198

Lam, T.Y.T., Cheung, M.F.K., Munro, Y.L., Lim, K.M., Shung, D., Sung, J.J.Y., 2022. Randomized Controlled Trials of Artificial Intelligence in Clinical Practice: Systematic Review. J Med Internet Res 24, e37188. https://doi.org/10.2196/37188

Langlotz, C.P., 2019. Will Artificial Intelligence Replace Radiologists? Radiol Artif Intell 1, e190058. https://doi.org/10.1148/ryai.2019190058

Lehman, E., Lialin, V., Legaspi, K.E., Sy, A.J., Pile, P.T., Alberto, N.R., Ragasa, R.R., Puyat, C.V., Taliño, M.K., Alberto, I.R., Alfonso, P.G., Moukheiber, D., Wallace, B., Rumshisky, A., Liang, J., Raghavan, P., Celi, L.A., Szolovits, P., 2022. Learning to Ask Like a Physician, in: Proceedings of the 4th Clinical Natural Language Processing Workshop. Presented at the ClinicalNLP 2022, Association for Computational Linguistics, Seattle, WA, pp. 74–86. https://doi.org/10.18653/v1/2022.clinicalnlp-1.8

Levine, D.M., Tuwani, R., Kompa, B., Varma, A., Finlayson, S.G., Mehrotra, A., Beam, A., 2023. The Diagnostic and Triage Accuracy of the GPT-3 Artificial Intelligence Model. https://doi.org/10.1101/2023.01.30.23285067

Li, Y., Wehbe, R.M., Ahmad, F.S., Wang, H., Luo, Y., 2023. A comparative study of pretrained language models for long clinical text. J Am Med Inform Assoc 30, 340–347. https://doi.org/10.1093/jamia/ocac225

Liebrenz, M., Schleifer, R., Buadze, A., Bhugra, D., Smith, A., 2023. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. Lancet Digit Health 5, e105–e106. https://doi.org/10.1016/S2589-7500(23)00019-5

Liu, X., Faes, L., Kale, A.U., Wagner, S.K., Fu, D.J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J.R., Schmid, M.K., Balaskas, K., Topol, E.J., Bachmann, L.M., Keane, P.A., Denniston, A.K., 2019. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. The Lancet Digital Health 1, e271–e297. https://doi.org/10.1016/S2589-7500(19)30123-2

Logé, C., Ross, E., Dadey, D.Y.A., Jain, S., Saporta, A., Ng, A.Y., Rajpurkar, P., 2021. Q-Pain: A Question Answering Dataset to Measure Social Bias in Pain Management. https://doi.org/10.48550/arXiv.2108.01764

Matheny, M., Israni, S.T., Ahmed, M., Whicher, D. (Eds.), 2019. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril.

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S., 2019. Dissecting racial bias in an algorithm used to manage the health of populations. Science 366, 447–453. https://doi.org/10.1126/science.aax2342

Ötleş, E., James, C.A., Lomis, K.D., Woolliscroft, J.O., 2022. Teaching artificial intelligence as a fundamental toolset of medicine. Cell Rep Med 3, 100824. https://doi.org/10.1016/j.xcrm.2022.100824

Palmer, K., 2023. How medical schools are missing the mark on artificial intelligence. STAT. URL https://www.statnews.com/2023/01/12/medical-school-artificial-intelligence-health-curriculum/ (accessed 2.3.23).

Plana, D., Shung, D.L., Grimshaw, A.A., Saraf, A., Sung, J.J.Y., Kann, B.H., 2022. Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review. JAMA Netw Open 5, e2233946. https://doi.org/10.1001/jamanetworkopen.2022.33946

Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S., Peng, L., Webster, D.R., 2018. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng 2, 158–164. https://doi.org/10.1038/s41551-018-0195-0

Rajkomar, A., Kannan, A., Chen, K., Vardoulakis, L., Chou, K., Cui, C., Dean, J., 2019. Automatically Charting Symptoms From Patient-Physician Conversations Using Machine Learning. JAMA Intern Med 179, 836–838. https://doi.org/10.1001/jamainternmed.2018.8558

Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Liu, P.J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G.E., Irvine, J., Le, Q., Litsch, K., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboum, S.L., Chou, K., Pearson, M., Madabushi, S., Shah, N.H., Butte, A.J., Howell, M.D., Cui, C., Corrado, G.S., Dean, J., 2018. Scalable and accurate deep learning with electronic health records. npj Digital Medicine 1, 1–10. https://doi.org/10.1038/s41746-018-0029-1

Rajpurkar, P., Irvin, J., Ball, R.L., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C.P., Patel, B.N., Yeom, K.W., Shpanskaya, K., Blankenberg, F.G., Seekins, J., Amrhein, T.J., Mong, D.A., Halabi, S.S., Zucker, E.J., Ng, A.Y., Lungren, M.P., 2018. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt

algorithm to practicing radiologists. PLoS Med 15, e1002686. https://doi.org/10.1371/journal.pmed.1002686
Ross, C., 2022a. How a complex web of businesses turned private health records from GE into a lucrative portrait of patients. STAT. URL https://www.statnews.com/2022/05/23/hipaa-patient-ge-data-privacy-profit/ (accessed 2.4.23).
Ross, C., 2022b. Microsoft's note-taking tool asks doctors to make a grand bargain: Share patient data, and get your life back. STAT. URL https://www.statnews.com/2022/11/07/microsofts-note-taking-tool-asks-doctors-to-make-a-grand-bargain-share-patient-data-and-get-your-life-back/ (accessed 2.28.23).
Sarraju, A., Bruemmer, D., Van Iterson, E., Cho, L., Rodriguez, F., Laffin, L., 2023. Appropriateness of Cardiovascular Disease Prevention Recommendations Obtained From a Popular Online Chat-Based Artificial Intelligence Model. JAMA. https://doi.org/10.1001/jama.2023.1044
Shah, N.R., 2019. Health Care in 2030: Will Artificial Intelligence Replace Physicians? Ann Intern Med 170, 407–408. https://doi.org/10.7326/M19-0344
Sheng, E., Chang, K.-W., Natarajan, P., Peng, N., 2019. The Woman Worked as a Babysitter: On Biases in Language Generation, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Presented at the EMNLP-IJCNLP 2019, Association for Computational Linguistics, Hong Kong, China, pp. 3407–3412. https://doi.org/10.18653/v1/D19-1339
Shortliffe, E.H., 2019. Artificial Intelligence in Medicine: Weighing the Accomplishments, Hype, and Promise. Yearb Med Inform 28, 257–262. https://doi.org/10.1055/s-0039-1677891
Ting, D.S.W., Cheung, C.Y.-L., Lim, G., Tan, G.S.W., Quang, N.D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I.Y., Lee, S.Y., Wong, E.Y.M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N.C., Finkelstein, E.A., Lamoureux, E.L., Wong, I.Y., Bressler, N.M., Sivaprasad, S., Varma, R., Jonas, J.B., He, M.G., Cheng, C.-Y., Cheung, G.C.M., Aung, T., Hsu, W., Lee, M.L., Wong, T.Y., 2017. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA 318, 2211–2223. https://doi.org/10.1001/jama.2017.18152
Topol, E., 2019. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, Illustrated Edition. ed. Basic Books, New York.
Tschandl, P., Rosendahl, C., Akay, B.N., Argenziano, G., Blum, A., Braun, R.P., Cabo, H., Gourhant, J.-Y., Kreusch, J., Lallas, A., Lapins, J., Marghoob, A., Menzies, S., Neuber, N.M., Paoli, J., Rabinovitz, H.S., Rinner, C., Scope, A., Soyer, H.P., Sinz, C., Thomas, L., Zalaudek, I., Kittler, H., 2019. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. JAMA Dermatol 155, 58–65. https://doi.org/10.1001/jamadermatol.2018.4378
Verghese, A., Shah, N.H., Harrington, R.A., 2018. What This Computer Needs Is a Physician: Humanism and Artificial Intelligence. JAMA 319, 19–20. https://doi.org/10.1001/jama.2017.19198
Veta, M., Heng, Y.J., Stathonikos, N., Bejnordi, B.E., Beca, F., Wollmann, T., Rohr, K., Shah, M.A., Wang, D., Rousson, M., Hedlund, M., Tellez, D., Ciompi, F., Zerhouni, E., Lanyi, D., Viana, M., Kovalev, V., Liauchuk, V., Phoulady, H.A., Qaiser, T., Graham, S., Rajpoot, N.,

Sjöblom, E., Molin, J., Paeng, K., Hwang, S., Park, S., Jia, Z., Chang, E.I.-C., Xu, Y., Beck, A.H., van Diest, P.J., Pluim, J.P.W., 2019. Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge. Med Image Anal 54, 111–121. https://doi.org/10.1016/j.media.2019.02.012

Wiggers, K., 2020. Google, Apple, and others show large language models trained on public data expose personal information | VentureBeat [WWW Document]. Venturebeat. URL https://venturebeat.com/2020/12/16/google-apple-and-others-show-large-language-models-trained-on-public-data-expose-personal-information/ (accessed 3.18.21).

Wu, S., Liu, S., Wang, Y., Timmons, T., Uppili, H., Bedrick, S., Hersh, W., Liu, H., 2017. Intrainstitutional EHR collections for patient-level information retrieval. Journal of the Association for Information Science and Technology 68, 2636–2648. https://doi.org/10.1002/asi.23884

Yang, X., Chen, A., PourNejatian, N., Shin, H.C., Smith, K.E., Parisien, C., Compas, C., Martin, C., Flores, M.G., Zhang, Y., Magoc, T., Harle, C.A., Lipori, G., Mitchell, D.A., Hogan, W.R., Shenkman, E.A., Bian, J., Wu, Y., 2022. GatorTron: A Large Clinical Language Model to Unlock Patient Information from Unstructured Electronic Health Records. https://doi.org/10.48550/arXiv.2203.03540

Zhou, Q., Chen, Z.-H., Cao, Y.-H., Peng, S., 2021. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools: a systematic review. NPJ Digit Med 4, 154. https://doi.org/10.1038/s41746-021-00524-2

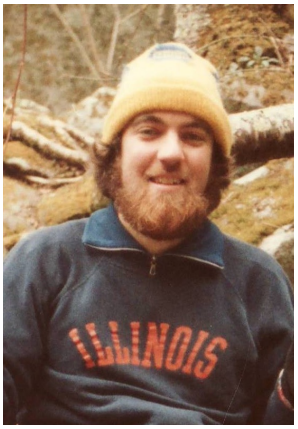
Translational Artificial Intelligence (AI): The Need to Translate from Basic Science to Clinical Value

William Hersh
Professor
Department of Medical Informatics & Clinical Epidemiology
Oregon Health & Science University
March 9, 2023 – PDF of slides and references at www.billhersh.info or from @williamhersh

1

Competing and other interests

- Funding for EHR cohort discovery research from
  - Alnylam Pharmaceuticals
  - PTC Pharmaceuticals
- Thanks to the 35 who have completed and the 7 currently enrolled in the 10x10 course



2

Translational AI – outline

- Promise of artificial intelligence (AI) and machine learning (ML) in medicine
- Current state of clinical impact of AI
- Research aiming to diagnose rare disease using ML
- Educating clinicians and informaticians
- Future directions

3

One-slide history of AI and ML in medicine

- A major activity of clinical informatics has been application of AI with aim of improving patient care (Shortliffe, 2019)
- First generation in 1960s-1980s
  - Focus on hand-crafted knowledge bases
  - Computers lacking data, processing power, GUIs, Internet, etc.
  - Led to "AI winter" in late 1980s and beyond
- Resurgence in 21st century
  - Driven by advances in ML, especially deep learning
  - Based on plentiful data, computer power, and networks
  - Overviews – Topol (2019), NAM (2019), Cohen (2022)
  - Still modest impact (as of 2023) in clinical settings

4

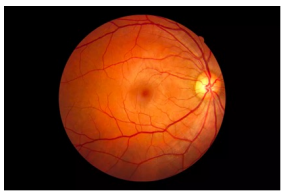
Promise of ML and AI in medicine

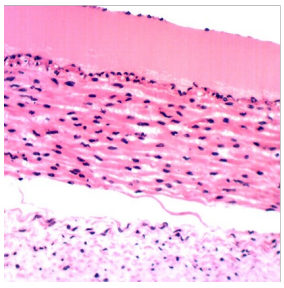
- Imaging
- Other applications
- Systematic reviews
- Building the evidence base

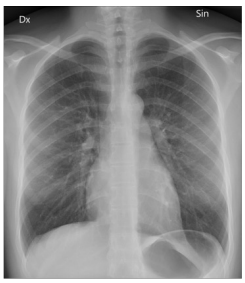
5

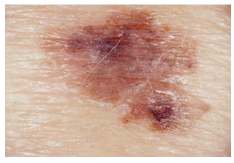
Imaging

- Earliest success for ML and AI
  - Diabetic retinopathy (DR) (Gulshan, 2016; Ting, 2017)
  - Histology of cancer (Bejnordi, 2017) and metastases (Veta, 2019)
  - Tuberculosis (Lakhani, 2017) and pneumonia (Rajpurkar, 2018)
  - Skin cancer (Esteva, 2017; Haenssle, 2018; Tschandi, 2019)
- Systematic review (Liu, 2019)
- State of the art (Esteva, 2021)





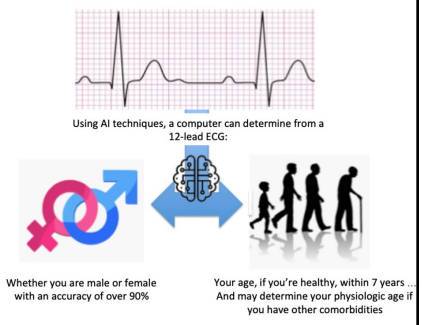




6

Many, many other application areas

- Wave forms – use of ECGs to determine age and sex (Attia, 2019)
- Age and sex determination from retinal images (Poplin, 2018)
- Length of stay, mortality, readmission, and diagnosis at two large medical centers (Rajkomar, 2018)
- Automatically charting symptoms from patient-physician conversations (Rajkomar, 2019)
- "Weakly supervised" (using clinical diagnoses) interpretation of pathology slides would allow pathologists to exclude 65–75% of slides while retaining 100% sensitivity (Campanella, 2019)
- AI system helped physicians extract relevant patient information from EHR in shorter time while maintaining high accuracy (Chi, 2021)



7

And now, large language models (LLMs)

- Building large models based on general and/or clinical text
- BERT models have transformed natural language processing capabilities, including in medicine, e.g., (Yang, 2022; Lehman, 2023; Li, 2023)
- And of course, ChatGPT has captured the public imagination, with some early successes in medicine
  - Diagnostic and triage accuracy for 45 vignettes comparable to physicians (Levine, 2023)
  - Answers to 21/25 questions about cardiovascular disease prevention deemed acceptable for patient-facing information platform and as AI-generated draft responses to questions sent by patients (Sarraju, 2023)
  - Performed at or near passing for three levels of USMLE (Kung, 2023)
  - My own experience (documented in my blog): some correct answers but also confabulation and not admitting it does not know

8

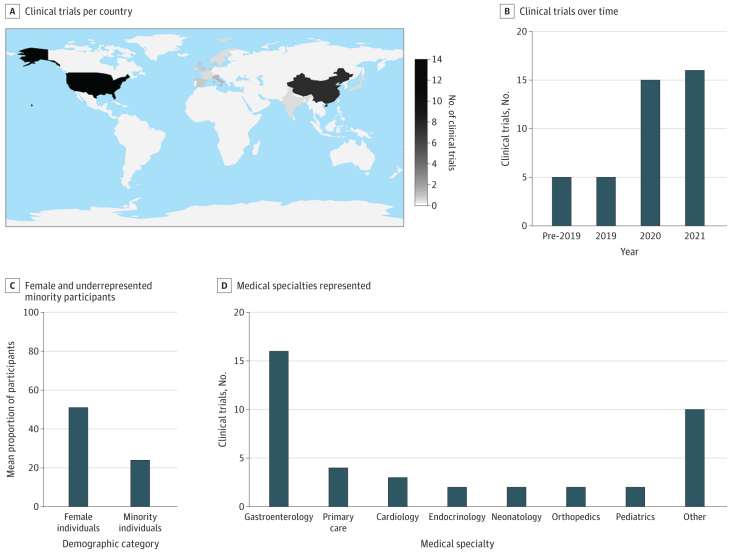
Current state of impact of AI

- From pubmed.gov
  - Tens of thousands of studies applying ML or AI
  - Hundreds of systematic reviews of ML and AI studies – mostly of models applied to clinical topics
- How many studies assessing ML/AI interventions using gold standard, RCT?
  - Collated in systematic reviews – earlier ones did not assess as rigorously, e.g., (Zhou, 2021; Lam, 2022)
  - Most recent and rigorous (Plana, 2022)

Systematic review of RCTs of ML/AI in healthcare (Plana, 2022)

- Exhaustive search of literature databases to identify RCTs of ML/AI interventions through October, 2021
- Excluded studies of non-RCT design, absence of original data, and evaluation of nonclinical interventions
- Identified 41 RCTs for further analysis
- Analyzed RCT characteristics, including primary intervention, demographics, adherence to CONSORT-AI reporting guideline, and Cochrane risk of bias

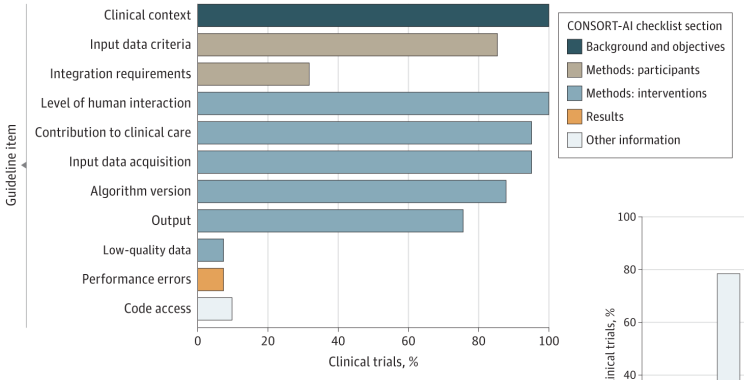
Where, when, who, and what (Plana, 2022)



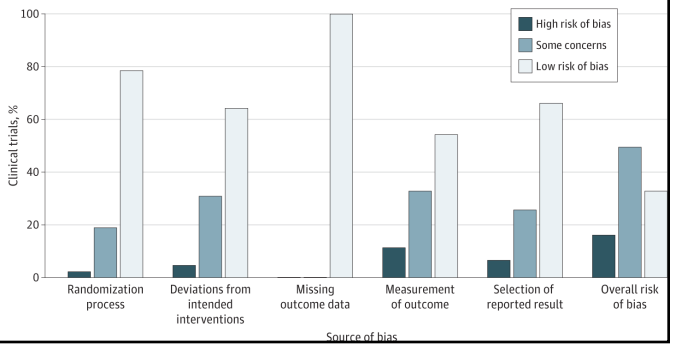
U of Iowa

11

Adherence to reporting guidelines and risk of bias (Plana, 2022)



- No assessment of study outcomes, unlike Zhou (2021) showing about 60% positive
- No attempted meta-analysis

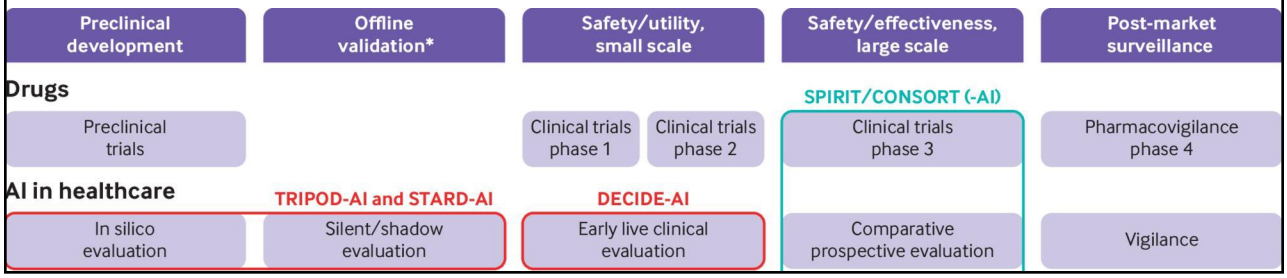


U of Iowa

12

Clear need to build evidence base

- As with all evidence-based medicine (EBM), major questions that guide patient care to ask about ML/AI include
  - Diagnosis – "test" for a "disease"
    - Ideally from comparison with "gold standard"
    - Predictive models are type of diagnostic test
  - Treatment – therapy or intervention to prevent or treat disease; ideally from randomized controlled trial (RCT)
- Studies of both questions can be aggregated into systematic reviews that may (if data allow) be combined via meta-analysis
- Appraisal tools being developed for AI studies (Ibrahim, 2021)



13

My work: applying information retrieval (search) to EHR data

- Use cases
  - Cohort discovery
  - Detection of rare diseases
- Data set
  - 100-300K records extracted from OHSU Research Data Warehouse (fully identifiable)
- Funded by grants from
  - NLM 1R01LM011934
  - Alnylam Pharmaceuticals
  - PTC Pharmaceuticals
- With help from OHSU collaborators
  - Steven Bedrick
  - Steven Chamberlin
  - Aaron Cohen
  - Tom Deloughery

(Hersh, 2020)

14

Cohort discovery

- Methods (Wu, 2017) and results (Chamberlin, 2020) for collection of 100K records
- R01 with Mayo and UT Houston renewed; updating data, systems, and methods
  - Standardizing on OMOP to share tools and methods across sites
- Major challenges
  - Heterogeneous records
  - Privacy concerns

Adults with IBD who haven't had GI surgery	Adults with inflammatory bowel disease who haven't had surgery involving the small intestine, colon, rectum, or anus.
Adults with a Vitamin D lab result	Adults with a lab result for 25-hydroxy Vitamin D collected between May 15 and October 15.
Postherpetic neuralgia treated with topical and systemic medication	Adults with postherpetic neuralgia ever treated by concurrent use of topical and non-opioid systemic medications.
Children seen in ED with oral pain	Children who were seen in the emergency department with herpetic gingivostomatitis, herpangina or hand, foot, and mouth disease, tonsillitis, gingivitis, or ulceration (aphthae, stomatitis, or mucositis) not due to chemotherapy or radiation.
3rd trimester prenatal visit with midwife or Ob/Gyn	Women who had a pregnancy with a 3rd trimester outpatient prenatal visit with an obstetrician and gynecologist or midwife.

Rare disease detection

- Over 1200 known rare disorders that affect < 1 in 200K patients worldwide, many under-diagnosed (https://rarediseases.org/; Haendel, 2020)
- Acute Intermittent Porphyria (AIP, aka Acute Hepatic Porphyria)
  - Rare genetic disease of heme biosynthesis – variable penetrance
  - Incidence 1 per 100K in population
  - Often undiagnosed for long time
  - Significant morbidity and effect on quality of life
  - "Neurovisceral" symptoms common with other diseases
    - Abdominal pain
    - Nausea and vomiting
    - Psychiatric changes
  - Diagnosed by inexpensive urine porphobilinogen test
  - New highly effective (and highly expensive) treatment available – RNA-silencing molecule givosiran (Balwani, 2020)

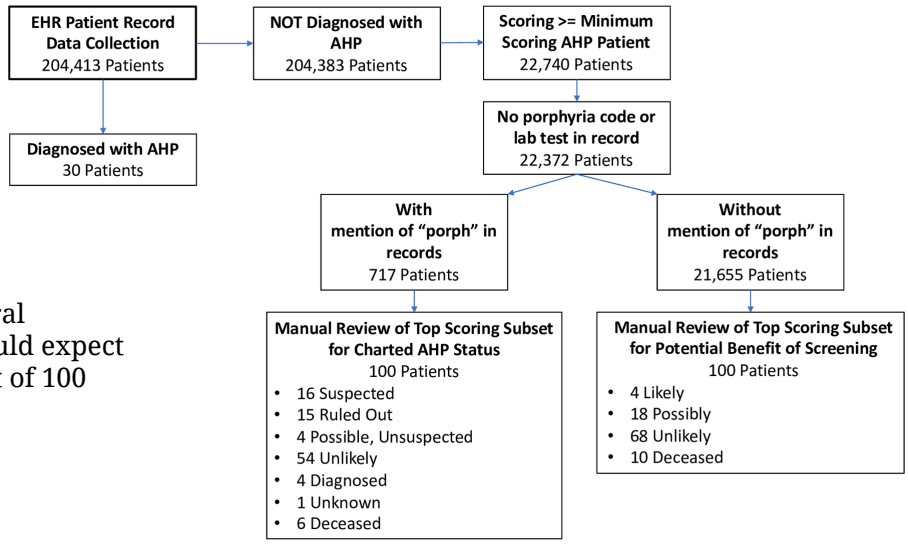
Can we detect rare diseases earlier using population-based techniques with EHR data?

- Funding from Alnylam Pharmaceuticals
- Expanded EHR data set to 200+ K patients
  - Updated base data set to 200K patients
    - Including from post-2015 era of ICD-10-CM coding
  - Enriched with 5,571 additional patients having "porph" in diagnoses, lab tests, and notes
- Preparation for machine learning
  - Positive training cases from ICD-10-CM E80.21 (47) with manual review to verify (30)
  - Negative training cases were the rest

ML approach (Cohen, 2020)

- Parsed EHR record into features – scored by frequency of appearance, labeled features by the EHR source document
- Univariate feature analysis – manually choose features not directly tied to provider attributes or suspecting patient had porphyria
  - e.g., "DeLoughery" and "cimetidine"
- Trained on full dataset, with best performance using support vector machine (SVM) with radial basis function (RBF) kernel
- Applied trained model back to full data set – ranked patients by margin distance

Aimed to identify patients with symptoms but no consideration of diagnosis of AIP

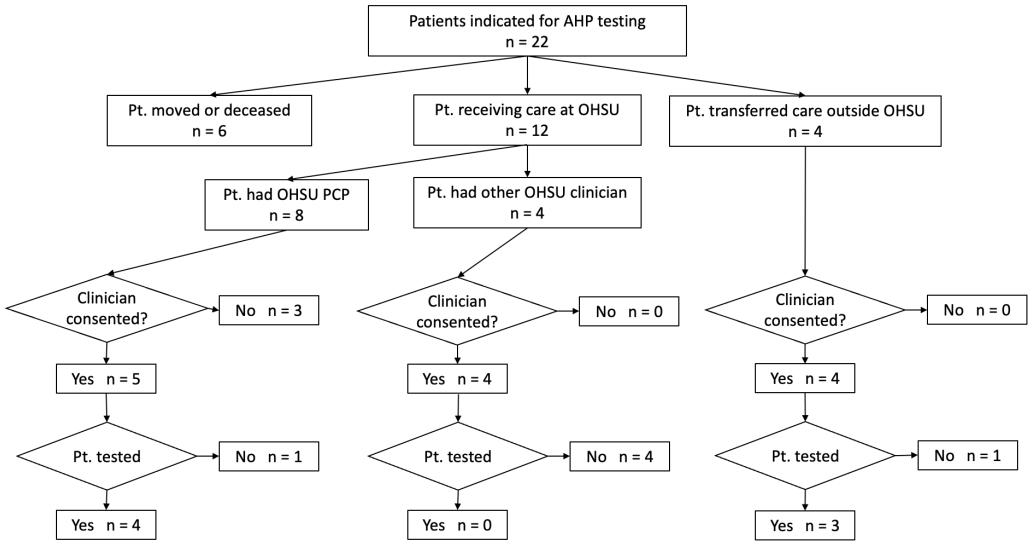


Note with natural prevalence, would expect 0.0005 cases out of 100

Clinical study (Hersh, 2022)

- Hindered by prolonged IRB process and COVID-19 pandemic, study was launched in late 2020
- IRB protocol required initial contact with primary care physician and, if they approved, offering the patient urine porphobilinogen testing
- Aimed to contact and enroll all 22 patients with AIP symptomatology but "unrecognized"

Clinician and patient participation



21

And the results showed…

- All 7 patients who came for testing had normal urine porphobilinogen
- Lessons learned
  - Clinical validation of machine learning models essential
  - Two-step approval required for patients not under our care but complicated
  - Rare diseases are rare
  - For other diseases, testing may be expensive and/or harmful
- Next steps
  - Discovery of patients with AADC deficiency (PTC Pharmaceuticals)

22

Other aspects of translational AI – clinicians must understand enough to use in practice

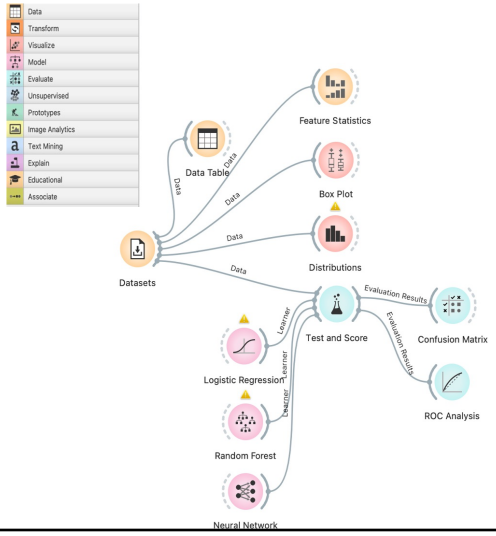
- Another competency in clinical informatics (Hersh, 2014; Hersh, 2022)
- Must be prepared for a clinical world influenced by AI (James, 2022)
- Medical schools may be "missing the mark" on AI (Palmer, 2023)
- AI should be taught as a "fundamental toolset of medicine" (Ötleş, 2022)

1. Find, search, and apply knowledge-based information to patient care and other clinical tasks
2. Effectively read from, and write to, the electronic health record for patient care and other clinical activities
3. Use and guide implementation of clinical decision support (CDS)
4. Provide care using population health management approaches
5. Protect patient privacy and security
6. Use information technology to improve patient safety
7. Engage in quality measurement selection and improvement
8. Use health information exchange (HIE) to identify and access patient information across clinical settings
9. Engage patients to improve their health and care delivery though personal health records and patient portals
10. Maintain professionalism through use of information technology tools
11. Provide clinical care via telemedicine and refer patients as indicated
12. Apply personalized/precision medicine
13. Participate in practice-based clinical and translational research
14. Apply machine learning applications in clinical care

23

Informaticians must be able to implement and evaluate

- Have implemented "applied" machine learning course at OHSU for those with modest math and programming backgrounds
- Role for visual programming tools, e.g., Orange?
  - (Hoyt, 2022; Hoyt, 2022)
  - https://orangedatamining.com/



24

Also cannot forget concerns about AI

- Data and algorithm bias
  - Clinical data may not be complete or accurate (Hersh, 2013)
  - Bias of clinicians and others may be "baked in" to data, leading to biased algorithms (Obermeyer, 2019; DeCamp, 2020)
- Language bias
  - Google searches for "professional" vs. "unprofessional" hair styles reveal racial differences (Alexander, 2016)
  - Training on large amounts of language "learns" biases inherent in text (Sheng, 2019; Logé, 2021)
- Privacy
  - Google, Apple, and others show large language models trained on public data expose personal information (Carlini, 2020; Wiggers, 2020)
  - May be compromised by need for data to be used to improve tools, e.g., Dragon voice recognition (Ross, 2022) and data re-identification (Ross, 2022)
- "Non-human" authors of scientific papers (Liebrenz, 2023; Flanagin, 2023)

How will ML and AI impact healthcare?

- Physicians (Jha, 2016; Jha, 2018; Shah, 2019) and ML (Verghese, 2018) must adapt
- "AI won't replace radiologists, but radiologists who use AI will replace radiologists who don't," (Langlotz, 2019)
  - True for all physicians, even Dr. McCoy?

Future directions

- As in all of medicine, results of basic science advances in AI must achieve clinical translation
- We must be cognizant of all types of bias and ensure fair and trustworthy AI
- As will all informatics, people are more important than technology
- Predictive models are important but must implement and evaluate in clinical settings
- *Great opportunity for informatics research and researchers!*

Thank you!

William Hersh, MD
Professor
Department of Medical Informatics & Clinical Epidemiology
School of Medicine
Oregon Health & Science University
Portland, OR, USA
http://www.ohsu.edu/informatics

Email: hersh@ohsu.edu
Web: www.billhersh.info
Blog: http://informaticsprofessor.blogspot.com
Twitter: @williamhersh