Computational and Structural Biotechnology Journal 21 (2023) 4536–4539

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Short communication

Multimodal analysis and the oncology patient: Creating a hospital system for integrated diagnostics and discovery

Christina Messiou [a,1], Richard Lee [b,1], Manuel Salto-Tellez [c,d,1,*]

[a] Imaging and Data Science Theme lead and Director of the Imaging AI hub at The Royal Marsden and Institute of Cancer Research, National Institute for Health Research Biomedical Research Centre, Sutton SM2 5PT, UK
[b] Consultant Respiratory Physician & Champion for Early Diagnosis Early Diagnosis and Detection Centre, NIHR Biomedical Research Centre at the Royal Marsden and ICR, National Heart and Lung Institute, Imperial College London, UK
[c] The Integrated Pathology Unit, the Institute of Cancer Research & The Royal Marsden Hospital, Sutton SM2 5PT, UK
[d] Precision Medicine Centre, Queen's University Belfast, UK

ARTICLE INFO

Keywords:
Multimodal Analysis
Integrated Diagnostics and Discovery
Computational Oncology
Radiomics
Digital Pathology
Artificial Intelligence
Health data

ABSTRACT

We propose that an information technology and computational framework that would unify access to hospital digital information silos, and enable integration of this information using machine learning methods, would bring a new paradigm to patient management and research. This is the core principle of Integrated Diagnostics (ID): *the amalgamation of multiple analytical modalities, with evolved information technology, applied to a defined patient cohort, and resulting in a synergistic effect in the clinical value of the individual diagnostic tools*. This has the potential to transform the practice of personalized oncology at a time at which it is very much needed.

In this article we present different models from the literature that contribute to the vision of ID and we provide published exemplars of ID tools. We briefly describe ongoing efforts within a universal healthcare system to create national clinical datasets. Following this, we argue the case to create "hospital units" to leverage this multi-modal analysis, data integration and holistic clinical decision-making. Finally, we describe the joint model created in our institutions.

The integration of clinical information for diagnostic and therapeutic decision-making is routine practice in oncology. The dominance of Multidisciplinary Teams, and Molecular Tumor Boards to achieve consensus on treatment [1] or efforts to create Specialist Integrated Hematological Malignancy Services [2] are clear examples. These transformative developments, however, have been hampered by a lack of a purpose-built digital frameworks to enable this cooperative effort, and rely on intuitive (subjective) coordination of the existing information to generate a clinical opinion or decide on a care pathway. The diversity, complexity, scale and pace of the data which informs patient management and scientific discovery have outgrown human capabilities and legacy IT systems.

There persists a desire to digitize information (and, in many instances, diagnostic visual materials such as radiology or tissue pathology) in reference hospitals. This is traditionally achieved with specialty-specific, purpose-built, silo-type systems, such as electronic patient records, Picture Archiving and Communication Systems (PACS) or laboratory information management systems. We suggest that an IT framework that would specifically unify access to hospital digital information silos, and enable integration of this information using machine learning methods, would bring a new paradigm to patient management and research. This is the core principle of Integrated Diagnostics (ID): *the amalgamation of multiple analytical modalities, with evolved information technology, applied to a defined patient cohort, and resulting in a synergistic effect in the clinical value of the individual diagnostic tools*. This has the potential of transforming, the practice of personalized oncology at a time in which this is very much needed: the number of patients who eventually benefit from genome-targeted therapy, even in environments in which genomic testing is offered across the board to cancer patients, is limited [3], whilst the performance of artificial intelligence has not reached yet the level that enables routine clinical applicability [4]. Meanwhile, novel therapies that are changing the landscape of oncology treatment, such as those targeting cancer immunity or homologous recombination deficiency pathways, lack robust

* Corresponding author at: The Integrated Pathology Unit, the Institute of Cancer Research & The Royal Marsden Hospital, Sutton SM2 5PT, UK.
E-mail addresses: richard.lee@rmh.nhs.uk (R. Lee), manuel.salto-tellez@icr.ac.uk (M. Salto-Tellez).
[1] All authors contributed equally to this work

https://doi.org/10.1016/j.csbj.2023.09.014
Received 7 August 2023; Received in revised form 13 September 2023; Accepted 13 September 2023
Available online 15 September 2023

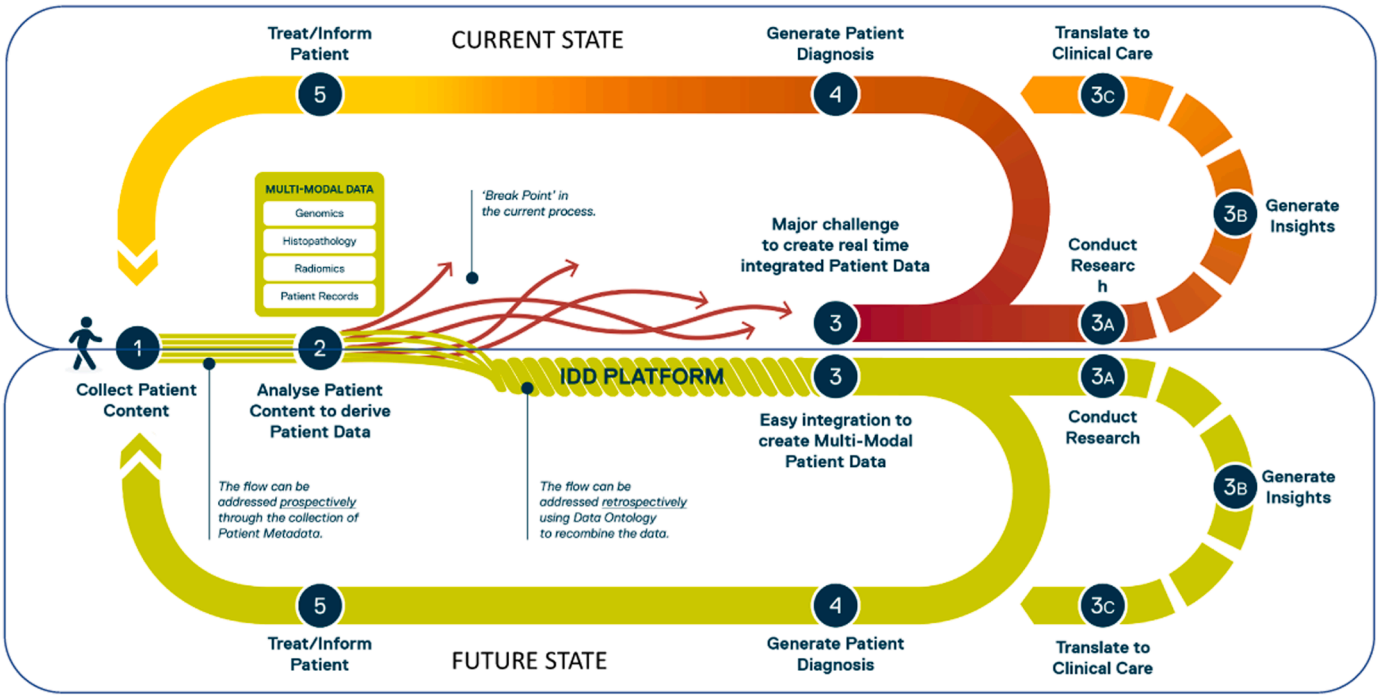

Fig. 1. Integrated diagnostic and discovery data model – the Gap. In contrast to the current state (red/orange), IDD (green) provides the capability and architectures for integration of multimodal data, generation of new insights and a feedback loop for clinical translation and ongoing learning.

biomarkers for prediction of therapeutic response. While some institutions are already applying AI algorithms for text analysis and subsequent clinical decision support in the context of known standards such as Fast Healthcare Interoperability Resources (FHIR), the generation and delivery of ID tools as described above is in its infancy and require more sophisticated IT interfaces.

Several models explain multi-modal analysis in the literature [5–9]. They all aim to fill a critical gap in the flow of information in hospital-based systems (see Fig. 1). They originate from different perspectives (e.g., discovery versus clinical implementation; laboratory medicine versus diagnostic imaging), but all have a common core structure:

- Multiple data modalities: the main "silos of information", including structured and discursive patient records, radiology, tissue pathology, genomic testing and (in general) laboratory medicine; but also endoscopy, the microbiome, information from biosensors, spatial transcriptomics or, more broadly, "multi-omics" analysis.
- A computational e.g., machine learning approach to data integration, including in-silico fusion of information at different levels or supervision: (a) strongly supervised methods such as such as random forest (RF), support-vector machine (SVM), or multilayer perceptron (MLP); (b) weakly supervised methods which include graph convolutional networks, multiple-instance learning and vision transformers; and (c) and fully unsupervised methods [5]. For the diagnostician using these tools, the approach to integration may have an effect in the hierarchical importance of the independent information silos to be integrated within the analytical tool, and thus in the final diagnostic performance of the new algorithm.
- A clear opportunity for patient benefit and a new frontier in precision health.

These approaches are beginning to deliver clear examples that, if developed further, may inaugurate more clear-cut solutions for patients e.g. improving current models for prediction of recurrence in breast cancer [10], delivering new prognostic and predictive models in targeted systemic anti-cancer therapy and immune-oncology [11], and improving pre-operative grading accuracy of soft tissue sarcoma by combining diagnostic imaging and tissue pathology [12]. Such combinatorial approaches, including those between images and free-text radiology and pathology reports can enable representation learning by drawing conclusions between linked data items, that could allow automated data annotation, which remains a significant barrier to AI approaches [13]. In general, some systems are allowing the "aggregation" of phenotypic, genotypic, financial and patient survey data into a single query-able model using the FHIR standard, but direct access-to-source and ML data integration still remains a challenge.

As in any other emerging area of medicine, differences in conceptual definitions will have an impact in our understanding of an "Integrated Diagnostic" tool. For example, Beauchamp et al. [14] consider ID as the process "wherein diagnostic data, together with clinical data from the electronic health record, are aggregated and contextualized by informatics tools to direct clinical action". This looser aggregation and consolidation by informatics tools, as opposed to the specific machine learning nature of the IDD definition we propose here, allows the authors to describe 24 ID tools, across all types of diseases, with 9 in the area of oncology. Once again, a general consensus of the scope of the ID concept may be necessary in the future to consolidate advancements in this new area of diagnostics.

Universal healthcare systems are beginning to establish national initiatives to curate clinical information of the highest possible quality for discovery purposes [15–18]. In the UK, there are several omniscient platforms with more clinical, or operational focus (such as the NHS Secure Data Environment and the proposed Federated Data Platform). These are acutely aware of concerns about data security and commercial exploitation, see Goldacre et al. [19]. More scientific repositories include the Data Alliance, UK Network, DARE, GECIP or the NCITA image repository. A good example of the latter is the CRUK sarcoma accelerator consortium database [20], developed at the Institute for Cancer Research in London, which uses the XNAT based infrastructure as the single point of data access for radiology, pathology, clinical data and multi-omics sourced from multiple collaborators [21]. This valuable global data asset in a rare cancer is enriched with curation and annotation tools including AI and a data dashboard for intuitive end-user

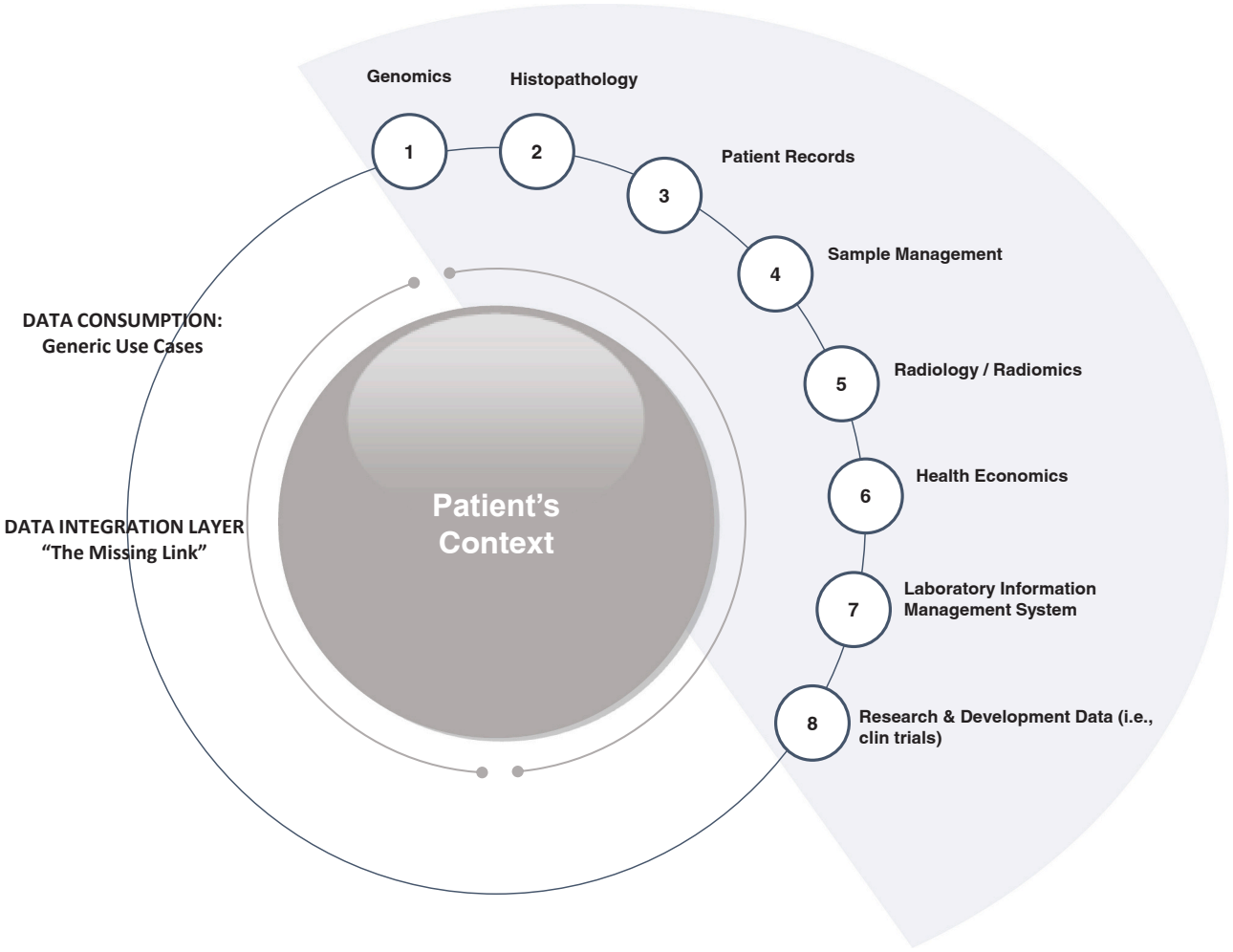

Fig. 2. Integrated Discovery and Diagnostic Data Model – the Components. The figure depicts the main silos or information that are commonly encountered in an academic hospital environment - genomics, histopathology, patient records, sample management, radiology/radiomics and laboratory information management systems. IDD also includes other areas of relevance such as R&D data and information relating to the cost of procedures (clinical or diagnostics) to allow IDD contribution to health economics.

interaction.

While national initiatives are pertinent to make the most of "minimum datasets", the models of conceptualization and implementation of ID in the literature so far, have considered, predominantly, either the "hospital unit" or the university as the center of this new development. There is opportunity in aligning the data requirements of a healthcare provider and biomedical university partner so that we are synergistically creating excellence in front-line care and discovery. At a time in which patients wish to have access to their own personal health records, ID also provides an opportunity to synthesize an accessible diagnostic report and to empower patients to become partners with health care professionals in managing their health [22].

The partnership between the Royal Marsden Hospital (RMH) and its academic partner, the Institute for Cancer Research (ICR) in London is committed to bring ID to the cutting-edge of oncology patient care. The driving force behind this narrative has been the coming together of multidisciplinary diagnosticians, scientists and clinicians, and the unified endorsement and engagement from the senior leadership of both organizations. Although IT and computational capabilities are key, we also recognize that attention to human factors and willingness to break down the silos will be critical to the success story of ID. Medical practitioners and scientists of both institutions have created the Integrated Discovery and Diagnostics Initiative (IDD): a joint strategic initiative aimed at accelerating oncology discovery and patient diagnostics through advanced data accessibility and analytics. Key to this solution is the provision of curated, annotated and integrated genomic, histologic, radiologic and clinical data to our world-leading scientists and clinicians. The guiding hypothesis is that integrated, multi-modal data will drive the next generation of oncology discovery and patient care, and that modern technology will enable this to be achieved at scale and speed. IDD has four main objectives:

1. Create a multi-modal platform to integrate digital hospital information (diagnostic images and metadata) from our 20,000 cancer patients a year, which is available to all ICR/RMH scientists and clinicians, technically excellent and financially sustainable
2. Develop, with patients, intuitive digital solutions to optimize oncology research and patient care, establishing a clinical research translation and validation pipeline with cyclical innovation
3. Develop a streamlined, transparent and robust information governance process to accelerate programmatic scale innovation on large-scale information management, with the patients' interest at its core
4. Develop strategic academic and commercial partnerships to accelerate the delivery of innovative solutions to cancer patients with patients' information at its core

To achieve the vision of IDD, scientists, clinicians and data experts from both institutions have described what they believe are the main

silos of information that IDD should start bringing together. The result of this exercise is depicted in Fig. 2. This model is based on the conceptualization and delivery of a single information management system that is able to a) provide easy, real-time access to all information silos in the hospital to both clinicians and scientists; b) allow the generation and implementation of new ML models able to create a new generation of "integrated biomarkers". These two advantages, accessibility and integrability, are at the core of the RMH/ICR IDD model. The model itself has novel aspects when compared with other ID systems published to date: firstly, the inclusion of information related to health economics as an important component of data integration; more importantly, the realization that this model allows the integration of experimental laboratory or research data with real-world, routine patient management information as a real-time data dashboard, such that enabling the access of scientists to this resource in the context of a robust Trusted Research Environment, probably represents one of the most efficient translational tools to generate, test and validate new discoveries. Once delivered as hospital-based units, these integrated systems will have the opportunity of creating a network of leading cancer hospitals and contribute minimum datasets to other national and international data initiatives.

In the field of oncology biomarker development, where there is a constant search for the next new, disruptive technology. Our institutions believe that the way we integrate the vast amount of digital information and diagnostic materials in leading centers of excellence, as hospital units, may provide us with a new generation of information-based, complex biomarkers, which will allow our patients to live longer and better lives.

CRediT authorship contribution statement

Salto-Tellez: Conceptualization, Writing – original draft, Writing-Review and editing, **Christina Messiou and Richard Lee**: Writing-Review and Editing,

Declaration of Competing Interest

CM receives funding from the Royal Marsden Cancer Charity and NIHR Biomedical Research Centre at The Royal Marsden and Institute of Cancer Research. Richard Lee reports funding indirectly related to this work including Royal Marsden Cancer charity, CRUK, Innovate UK (co-funded by Roche), RM Partners Cancer Alliance, NIHR, NHS England, and personal private practice. MST is a scientific advisor to Mindpeak and Sonrai Analytics, and has received honoraria recently from BMS, Roche, MSD and Incyte. None of these disclosures are related to this work.

Acknowledgment

The authors would like to thank Adrian Cottrell, Lewis Butler, Cecil Lynch and Veera Bukshi for their critical and constructive appraisal of this manuscript. This article represents independent research supported by the National Institute for Health and Care Research (NIHR) Biomedical Research Center at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, London. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

References

[1] Minsky BD. Multidisciplinary case teams: an approach to the future management of advanced colorectal cancer. Br J Cancer 1998;77(Suppl 2):1–4 (Suppl 2).
[2] Cartwright A, Snowden JA, Whitehouse H, Scott S, Whitby L. Implementation of the updated NICE haematological cancers (NG47) improving outcomes guidelines across specialist integrated haematological malignancy diagnostic services (SIHMDS) in England: a UK NEQAS LI survey. J Clin Pathol 2023;76(9):618–23.
[3] Marquart J, Chen EY, Prasad V. Estimation of the percentage of US patients with cancer who benefit from genome-driven oncology. JAMA Oncol 2018;4(8):1093–8.
[4] Echle A, Grabsch HI, Quirke P, van den Brandt PA, West NP, Hutchins GGA, et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. Gastroenterology 2020;159(4):1406–16.
[5] Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, et al. Artificial intelligence for multimodal data integration in oncology. Cancer Cell 2022;40(10):1095–110.
[6] He X, Liu X, Zuo F, Shi H, Jing J. Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. Semin Cancer Biol 2023;88:187–200.
[7] Lippi G, Plebani M. Integrated diagnostics: the future of laboratory medicine? Biochem Med 2020;30(1):010501.
[8] Cui C, Yang H, Wang Y, Zhao S, Asad Z, Coburn LA, et al. Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: a review. Prog Biomed Eng 2023;(2):5. https://doi.org/10.1088/2516-1091/acc2fe.
[9] Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med 2022;28(9):1773–84.
[10] Sammut SJ, Crispin-Ortuzar M, Chin SF, Provenzano E, Bardwell HA, Ma W, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature 2022;601(7894):623–9.
[11] Vanguri RS, Luo J, Aukerman AT, Egger JV, Fong CJ, Horvat N, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. Nat. Cancer 2022;3(10):1151–64 (Consortium; Hellmann MD, Sauter JL, Shah SP).
[12] Mcaddy NC, Hallin M, Strauss D, Smith M, Hayes A, Yusuf S, et al. CT imaging improves histopathological grading of retroperitoneal leiomyosarcomas. Eur J Surg Oncol 2020;46(2):288–92.
[13] Zhou HY, Chen X, Zhang Y, et al. Generalized radiograph representation learning via cross-supervision between images and free-text radiology reports. Nat Mach Intell 2022;4:32–40.
[14] Beauchamp NJ, Bryan RN, Bui MM, Krestin GP, McGinty GB, Meltzer CC, et al. Integrative diagnostics: the time is now-a report from the international society for strategic studies in radiology. Insights Imaging 2023;14(1):54.
[15] Ortega-Calvo Alberto S, Morcillo-Jimenez Roberto, Fernandez-Basso Carlos, Gutiérrez-Batista Karel, Vila Maria-Amparo, Martin-Bautista Maria J. AIMDP: An artificial intelligence modern data platform. Use case for Spanish national health service data silo. Future Gener Comput Syst 2023;143:248–64.
[16] ⟨https://www.hdruk.ac.uk/about-us/contact-us/our-locations/⟩; date accessed on 10th September 2023.
[17] Sebire NJ, Cake C, Morris AD. HDR UK supporting mobilising computable biomedical knowledge in the UK. BMJ Health Care Inf 2020;27(2):e100122.
[18] ⟨https://www.england.nhs.uk/digitaltechnology/digitising-connecting-and-transforming-health-and-care/⟩. Date accessed 10th September 2023.
[19] Goldacre B., Morley J., Hamilton N. et al. Better, broader, safer: using health data for research and analysis. ⟨https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis⟩ 2022. Date accessed: 5th May 2023.
[20] ⟨https://www.icr.ac.uk/our-research/centres-and-collaborations/the-royal-marsden-partnerships/joint-sarcoma-research-centre/sarcoma-accelerator-consortium⟩. Dare accessed 10th Septenber 2023.
[21] McAteer MA, O'Connor JP, Koh DM, Leung HY, Doran SJ, Jauregui-Osoro M, et al. Introduction to the National Cancer Imaging Translational Accelerator (NCITA): a UK-wide infrastructure for multicentre clinical translation of cancer imaging biomarkers. Br J Cancer 2021;125(11):1462–5.
[22] Silberner J. Access to records: do open notes work for patients. BMJ 2023;381:789. https://doi.org/10.1136/bmj.p789.