BMJ Health & Care Informatics

Software development skills for health data researchers

Caroline Morton ,[1] Nicholas Devito ,[1] Jessica Morley,[1] Iain Dillingham,[2] Anna Schultze ,[3] Sebastian Bacon ,[4] Peter Inglesby,[4] Steven Maude ,[1] Ben Goldacre[4]

To cite: Morton C, Devito N, Morley J, *et al*. Software development skills for health data researchers. *BMJ Health Care Inform* 2022;**29**:e100488. doi:10.1136/bmjhci-2021-100488

Accepted

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
[2]University of Oxford, Oxford, UK
[3]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK
[4]Primary Care Health Sciences, University of Oxford, Oxford, UK

Correspondence to
Dr Caroline Morton;
caroline.morton@phc.ox.ac.uk

INTRODUCTION

Health data researchers are increasingly required to develop complex analytic code in order to implement sophisticated analyses on large health datasets. While writing analysis scripts (box 1) for academic projects is distinct from general purpose software development, they share many of the same features. A researcher's script usually consists of a sequence of commands executed by a computer to extract, reshape, clean, describe and analyse data. If the quality of this analytic code cannot be reasonably assured, then results cannot be trusted: programming errors have resulted in high profile retractions.[1–3] Similarly, if lengthy scripts for data management cannot be re-used, then work is needlessly duplicated.

The software engineering community has developed a range of techniques to improve the quality, re-usability, efficiency and readability of code. Organisations such as the Software Sustainability Institute[4] support this approach to code development and provide more detailed guidance and education which are well worth reviewing. In this brief guide we explain how researchers can borrow best practices and freely available tools from this community to improve their work. We specifically cover the following three topics: Writing High Quality Code, Working Collaboratively and Sharing your work. Throughout the piece we often refer to examples from Python or R, two popular open source programming languages used by academics, but our advice is universal and there will be analogues to these examples in any commonly used statistical or general purpose programming language.

METHODS

In this section, we introduce the three major themes and break down each theme with some key concepts and practical guidance.

Writing high quality code

Writing high quality code goes beyond the complexities of the analytic script itself, and should include documentation on what the code does, what decisions were taken and where, and how to recreate the same scripting environment in which the code runs. It can also include introducing efficiencies by encapsulating repeated code into functions that can be reused by you and others. Many programming languages also have style standards and specific recommendations on how to format and construct your code, like PEP8 for Python[5] and the tidyverse for R[6]. While the specifics of these for any individual language are outside the scope of this article, it is worth looking into to make sure your code is readable and quickly understandable to others. Integrated Development Environments (IDEs) such as PyCharm and R Studio are software applications that can integrate the coding standards and highlight places in your code where these standards are violated. They also provide a number of other useful features that can help you work more effectively and efficiently such as syntax highlighting, code autocompletion, code search, and tools to find errors and run unit tests.

Documentation

Analytic scripts can be long and complex, and good documentation can improve reusability and understanding by providing information about what each section of the scripts is doing, and why. Increasing the readability of the code improves your user's understanding, increases the likelihood that other people will use your code, and acts as an aide memoire when you return to your work after a period of time.

How to write and share good documentation

The simplest form of documentation is as a "comment" in-line with the code: these are text notes embedded in the code, marked so

Box 1 Glossary

Analytic Script: A series of commands written in a programming or statistical language such as R, Stata or Python, that are executed by a computer. These commands are used to *do* the analysis and may involve data extraction, cleaning, processing and analysis.
Commit: An individual change or revision to a file or set of files[9]
Docstring: This is a non-executable text that is attached to units of code such as functions, and documents what the code is doing. For example, this may include inputs, outputs, and specific errors.
Functions: These are pieces of code that can be run (or invoked) and executes the code specified.
Library: This is a collection of code that does a particular task or set of tasks, and can be imported and used in other projects.
Open source: Code or software projects where the source code is freely available and may be changed, and shared by others.
Pull: This is the term that describes when you fetch files from GitHub or similar. You can "pull" the most up to date file onto your computer, or "pull" changes that your colleague may have made.[9]
Pull Request: There are proposed changes to a repository by a user and are accepted or rejected, or commented on by the other project collaborators.[9]
Push: This is the term that describes when you send your committed changes back to GitHub (or a similar platform). Once pushed, others will be able to see your suggested changes to any files.[9]
Repository: This is a project space within GitHub or GitLab that holds a project. The easiest way of conceptualising this is as a folder that contains all your project files, and stores each files' revision history.[9]
Requirements/Dependencies: These are software libraries that are required to run a particular project or piece of code. They normally have a version number, for example, version 0.0.1, 0.0.2 etc

as not to be executed, that provide plain-English context for what is occurring in the adjacent commands. If your code is complex, and you have converted repeating code patterns into "functions" (as discussed below), then you can also build more formal documentation attached to these functions; in Python, for example, these are called "docstrings". These are less like incidental comments for a few lines of code, and more like formal documentation that describes how a particular block of code can be invoked and used. Ideally, all code would also have some overarching contextual documentation. For researchers we recommend that this should include at minimum: simple instructions, including the order in which programmes should be run; project details (called a "readme" file); and a link to the research protocol. Ideally it should have enough information for a researcher to be able to "recreate" the software environment in which the research was run in (see below for more information on environments).

Cataloguing your environment

Analysis scripts and other forms of software do not exist in isolation: they are written to be executed in particular environments. A snapshot file, such as Python's "requirements.txt", captures those assumptions, to tell users the exact version of the programming language or statistical analysis packages (often called dependencies) that are needed for the code to execute. When executing code in a "walled garden" environment (such as the Stata software with no bespoke added libraries) it is sufficient to simply give the version number of the single piece of software used; in more complex environments, good cataloguing is vital. Software is constantly evolving and advancing; commands that once worked in a certain way may have changed their implementation, such that there are small or large differences in the output from a given command. By providing adequate information about the requirements of your code, someone else can accurately run your code.

How to Catalogue your environment

The exact name and process for creating a requirements file can vary by programming language but the idea is the same. Sometimes this documentation takes the form of a simple text file in your project repository that lists the software packages used, and their version numbers. This can be generated manually, but for complex environments and repeated use it is often better to automate cataloguing with tools such as "pip-tools" for Python. Other tools exist for more advanced users to create reproducible virtual environments or full virtual machines like Docker.[7 8]

Functions

It is common for the same task to be performed many times over in a given analysis, or across projects. Inexperienced coders will often copy and paste code "patterns", with minor changes, to perform repetitive tasks. More experienced programmers aim to replace these code patterns with reusable "functions", which group the repetitive tasks together into single units of code with their associated documentation. Using functions has the obvious benefit of reducing the risk of errors when having to make small changes to a part of the code, as the changes are made once within a function.

How to write a function

We use the term "function" here for simplicity: however the exact names and mechanisms for creating this kind of reusable code will vary by language and purpose (for example, "macros" in Stata are essentially the same as functions). All methods tend to share the same basic structure: creating generalisable code that takes defined inputs, executes, and then returns a standard output.

Unit tests

When repetitive tasks are grouped together into functions, these functions can be more easily "tested" to check that their observed behaviour matches their expected behaviour. Performing these checks manually is tedious and error-prone for humans, so programming languages provide additional tools to automate this process. Central to these tools are "unit tests": pieces of code that systematically test a "unit" of code such as a function. They provide the function to be tested with a range of controlled inputs and allow the programmer to make assertions about the

expected outputs, to verify that the function is performing as expected.

How to write a unit test

Tests are important: they allow you to change small parts of a complex analytic codebase confidently, with a safety net, knowing that many errors will be caught early. The programmer can run tests individually or in groups when writing code. There are also automatic integrations via platforms like GitHub or GitLab that run tests automatically each time new code is committed. It is a good idea to follow the "Arrange, Act, Assert" principle.[8] *Arrange* a suitable curated input for the function to be tested on: for example, if the function transforms data, then recreate a much smaller version of that dataset where the correct function output has been pre-calculated. Then *Act* by passing this pre-prepared test dataset to the function that is being tested, and record the answer. Lastly, *Assert*: compare the output you got from the tested function with your earlier calculation. This could be done manually or preferably via code to assert that these two outputs match each other.

Working collaboratively

Software engineers and health data researchers usually work in teams and need to collaborate effectively. Software engineers are well-versed in using tools such as GitHub for collaborative working, and these tools have a low barrier to entry for health-data researchers. In this section, we will introduce GitHub and how it can facilitate best practices of version control, and code review, within a team.

Using Github to share and manage code

All of the working practices described in this paper are supported by commonly used software tools, of which GitHub is the most prevalent. The key to good practice in software development is the use of a strong platform that facilitates iterative development with version control, code review, unit testing, and code sharing.[9] GitHub is a good option as it is freely available for both private and public projects, well documented and supported, and friendly to beginners; other good alternatives such as GitLab also exist.

How to get started with GitHub

Users can sign up via www.github.com and make free accounts. This gives unlimited space for projects called repositories. Research groups may benefit from more advanced functionality that do have some associated costs. Projects can be changed from private to public, and vice versa, so it is possible to develop your code in private, and then share on publication of the associated paper, if that is a preferred pipeline.

Version control

Version control is the process of tracking and managing a project's code throughout its development. Software platforms keep track of all changes made to the code and allow multiple researchers to work on the same code at the same time. Changes can then be merged back into one "main" codebase. Archives of these changes are automatically logged for future reference, with a record of who made each change; and changes to sections of code can be visualised for ease of comparison. It also provides a safety net, as code can easily be reverted back to an earlier version if a problem is encountered later on in the project.

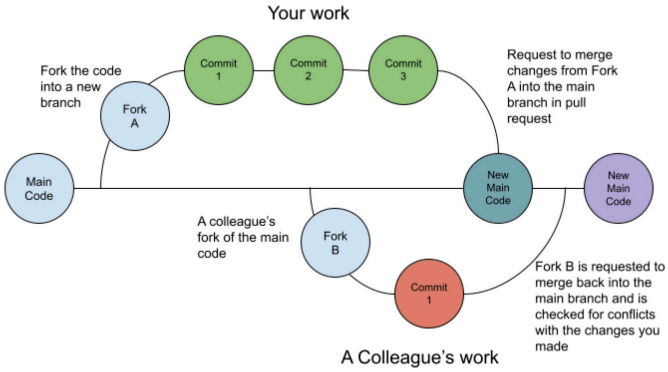

Figure 1 This figure shows an example workflow for a colleague and you using git to work on the same repository. In it, you fork your code (copy the repo) to work safely on the code whilst the current main branch remains untouched. You commit your changes and request to merge them back into the main branch. If accepted, these changes become part of the main code. Future merges by colleagues will be checked for conflicts since they were working on an earlier version of the code.

How to do version control

GitHub and other similar platforms facilitate version control as a built-in feature. Small changes to the code are submitted (called "commits") and tracked. During development you can "clone" a copy of the repository to safely work on while the current codebase remains untouched. While users are pointed to the stable main code "branch", you can safely revise, update, and experiment with your code until you are ready to commit the changes (figure 1).

Often you will propose changes to a repository in a "pull request" that documents all the edits you have made and are now proposing to be written over the canonical "main" version of the code. These pull requests act as a natural inflection point to ask for a code review (see below), and ensure none of your changes conflict with the current state of the repository. When a pull request is accepted and "merged" a history of all commits are maintained within the repository. This allows users to revisit any prior development state of the code, and provides transparency into the development of the project (figure 2).

Code review

Code will often contain shortcomings, or errors. A single incorrect character may have a catastrophic impact on the outputs of an analysis: in the recent past this has led to numerous retractions or corrections,[1 2] and it is likely

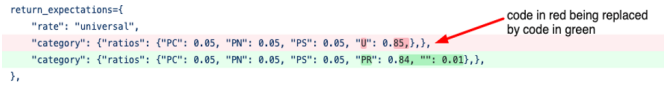

Figure 2 This screenshot of a pull request compares new code against existing code in the browser on GitHub. It shows proposed "new" code additions or edits in green, and code that is being removed or changed in red. Code that has not changed remains white.

that many coding errors go unnoticed. On some teams a single person may be responsible for writing all the code for a project. Code review typically involves a separate person examining the code, and sometimes running it, in order to spot issues. It aims to guard against error, and provide a useful opportunity for feedback or suggested amendments to improve the efficiency and readability of the code. Some research groups have implemented code reviews and have openly recommended this practice because of the benefits in quality and reproducibility.[10] We believe code review is essential and hard work and reviewers should be acknowledged as full members of the study team.

How to do code review

There is no one method for code review; however in general it is best to review often, and not at the end of the project, and for both the researcher and the code reviewer to have clear expectations of what code review will entail. For example, does it include running the code entirely or simply looking and commenting on the code. Google has produced some guidance[11] on how to think about and implement successful code review practices.

Some groups find it effective to use a '*buddy*' system where all code and outputs are checked by at least one other knowledgeable member of the study team for bugs and suggestions made for simple improvements. This can involve looking over a pull request, or checking an entire project to ensure it runs as expected. When you feel confident that your code does what is intended you can share it with the wider community which will ideally generate even more review and feedback. Code review is also one of the benefits of making code publicly available: having your code published enables other research teams as well as peer reviewers to assess the analytical code underlying any given study for accuracy. Even a cursory code review is better than none at all.

Code sharing

Sharing the code that underlies your analyses is a quick, cheap, and easy way to provide transparency into your methods. Your code can usually be shared without many of the concerns around privacy and disclosure that can complicate data sharing. Other researchers working in the field can re-use and learn from code, with credit, for their own projects: this increases the efficiency of research, and may open the door to new collaborations. In open source software development it is standard practice for others to offer suggestions, improvements, or entirely new features to existing repositories. Making your code available may be the first steps towards future collaborations and making a more generalizable tool for the wider research community.

How to share code

Code can be shared in a variety of ways: the simplest option is to share code in an appendix to a paper; however it is better to use one of the free software development platforms, such as GitLab or GitHub, which provide additional benefits and usability to interested users as discussed above. These services allow users to develop and share code in a "repository", which can be thought of as a project folder for each piece of work. In addition, users can interact with these platforms through simple graphical user interfaces, which is useful for those unfamiliar with working at the command line of an operating system. These platforms are indexed by major search engines meaning that your work is also more likely to be *found*. After uploading your code you can apply appropriate licenses that allow re-use of the software with or without restriction, modification, or citation. It is also easy to generate a digital object identifier (DOI) for specific versions of your code released through GitHub, by archiving through a service such as zenodo. GitHub also recently added support for citations files added directly to repositories.[12] In our view researchers should always cite other researchers' code when re-using it, or deriving insights from it: however as a formality we tend to use the MIT licence.[13]

Libraries

Useful functions, and their associated unit tests, often outgrow individual projects, and build a broader user-base. When they do, more experienced programmers move them into reusable code "libraries" and share them through package indexes or archive networks. By creating a library, researchers contribute to the broader research community. This more advanced variety of code sharing is common in many areas of scientific research, such as Geographic Information Science, but it is less common at present in health data research.[14–16]

How to create a library

Programming language communities have developed the tools to create and share code libraries easily through package indexes or archive networks. Python, for example, has PyPI, or the Python Package Index; R has CRAN, or the Comprehensive R Archive Network.

DISCUSSION

We hope this introduction into some of the basics of software development best practice is helpful to researchers of various levels of coding experience. Implementing the practices that fit your group's workflow can increase productivity, facilitate open collaboration with the larger community, and ultimately lead to higher quality research. Importantly, in other disciplines, sharing code with good

documentation has already been seen to produce quality and efficiency benefits for the wider research community.[17 18]

We recognise that there are barriers to embracing these practices: trying to do more with your code, beyond simply scripting out the analysis, can be intimidating; good code can be under-appreciated; and implementing these concepts in your own work may require familiarising yourself with new tools, jargon, and ways of thinking. A key area for development should be establishing communities of practice in research software to empower and educate researchers to use the tools that are available, in a way that works with their domain and team. Like-minded analysts with the UK NHS, for instance, have established an NHS R community to share knowledge, tools, and guidance among their peers.[19] Software Carpentry and Data Carpentry have sought to do the same by running an introductory course followed on by support to run monthly engagement to develop a local community of practice[[20]]. Senior leadership buy-in to the value of these communities has been key to getting them running. Online forums such as StackOverflow have been set up by software developers to allow people to ask questions about how to solve problems when writing and implementing their code. These contain a knowledge-base of thousands of answered questions covering a wide array of topics and domains with the ability to ask new questions if yours isn't covered.

Funders and journals may not fully appreciate that a well maintained and widely used open source library is as valuable as a high profile publication. We anticipate that research funders and leaders will increasingly recognise the value of software and its tools to the quality and efficiency of research.[21–23] Journals could consider mandating code sharing at the time of publication and even simple moves such as establishing a software policy for the journal would encourage code to be shared.

CONCLUSION

We strongly believe that researchers should aim to embrace modern best practice around software development because increasingly, in the era of data-driven research, research *is* software development. For this to occur, funders and journals need to buy-in to its value, and encourage individuals and teams to adopt the tools and techniques employed by the software development community.

Twitter Caroline Morton @dr_c_morton

Contributors This paper is written by members of the Datalab (University of Oxford) who worked with large datasets to create both traditional academic papers and interactive data-driven tools. We are a mixed team of software developers, clinicians and researchers, and many of us work across these domains, writing software to do research, and as such, we incorporate many of the modern software development techniques and tools described in this paper. CM, ND, JM and BG developed the idea for this paper. CM, ND and BG wrote the first draft and all other authors contributed to subsequent drafts.

Funding BG has received research funding from the Laura and John Arnold Foundation, the NHS National Institute for Health Research (NIHR), the NIHR School of Primary Care Research, the NIHR Oxford Biomedical Research Centre, the Mohn-Westlake Foundation, NIHR Applied Research Collaboration Oxford and Thames Valley, the Wellcome Trust, the Good Thinking Foundation, Health Data Research UK (HDRUK), the Health Foundation, and the WHO; he also receives personal income from speaking and writing for lay audiences on the misuse of science.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study.

ORCID iDs
Caroline Morton http://orcid.org/0000-0003-0783-0042
Nicholas Devito http://orcid.org/0000-0001-8286-1995
Anna Schultze http://orcid.org/0000-0002-1637-837X
Sebastian Bacon http://orcid.org/0000-0002-6354-3454
Steven Maude http://orcid.org/0000-0002-6112-5101

REFERENCES

1 Kang JH, Boumenna T, Stein JD, *et al*. Association of statin use and high serum cholesterol levels with risk of primary open-angle glaucoma. *JAMA Ophthalmol* 2019;137:756–65.
2 Gander JC, Zhang X, Ross K, *et al*. Notice of Retraction and Replacement. Gander et al. Association Between Dialysis Facility Ownership and Access to Kidney Transplantation. JAMA. 2019;322(10):957-973. *JAMA* 2020;323:1509–10.
3 Goldacre B, Morton CE, DeVito NJ. Why researchers should share their analytic code. *BMJ* 2019;367:l6365.
4 The software sustainability Institute. Available: https://www.software.ac.uk/ [Accessed 16 Nov 2021].
5 PEP 8 - Style Guide for Python Code. Available: https://www.python.org/dev/peps/pep-0008/ [Accessed 16 Nov 2021].
6 Tidyverse. Available: https://www.tidyverse.org/ [Accessed 16 Nov 2021].
7 Docker. Empowering APP development for developers. Available: https://www.docker.com/ [Accessed 22 Feb 2022].
8 12. virtual environments and packages — python 3.10.2 documentation. Available: https://docs.python.org/3/tutorial/venv.html [Accessed 22 Feb 2022].
9 Github. GitHub glossary. Available: https://docs.github.com/en/github/getting-started-with-github/github-glossary [Accessed 27 Apr 2021].
10 Vable AM, Diehl SF, Glymour MM. Code review as a simple trick to enhance reproducibility, accelerate learning, and improve the quality of your team's research. *Am J Epidemiol* 2021;190:2172–7.
11 How to do a code review. Available: https://google.github.io/eng-practices/review/reviewer/ [Accessed 26 Apr 2021].
12 Github. About citation files. Github. Available: https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-citation-files [Accessed 16 Nov 2021].
13 The MIT license. Available: https://opensource.org/licenses/MIT [Accessed 22 Feb 2022].
14 GeoPandas 0.9.0 — GeoPandas 0.9.0 documentation. Available: https://geopandas.org/ [Accessed 26 Apr 2021].
15 Lifelines — lifelines 0.25.11 documentation. Available: https://lifelines.readthedocs.io/en/latest/ [Accessed 26 Apr 2021].
16 PySAL. Available: http://pysal.org/ [Accessed 26 Apr 2021].
17 Wilson G, Bryan J, Cranston K, *et al*. Good enough practices in scientific computing. *PLoS Comput Biol* 2017;13:e1005510.
18 Science as amateur software development, 2020. Available: https://www.youtube.com/watch?v=zwRdO9_GGhY [Accessed 27 Apr 2021].
19 NHS-R Community - Promoting the use of R in the NHS, 2017. Available: https://nhsrcommunity.com/ [Accessed 16 Nov 2021].

20 Teal TK, Cranston KA, Lapp H, *et al*. Data Carpentry: workshops to increase data literacy for researchers. *Int J Digit Curation* 2015;10:135–43.
21 Knowles B. Introducing data for science and health at Wellcome. Wellcome data, 2021. Available: https://medium.com/wellcome-data/introducing-data-for-science-and-health-at-wellcome-de49b069e6b [Accessed 27 Apr 2021].
22 Chan Zuckerberg Initiative. Essential open source software for science (EOSS) -, 2019. Available: https://chanzuckerberg.com/eoss/ [Accessed 27 Apr 2021].
23 Department of Health and Social Care. Goldacre review. GOV.UK, 2021. Available: https://www.gov.uk/government/news/new-review-into-use-of-health-data-for-research-and-analysis [Accessed 27 Apr 2021].

BMJ Health & Care Informatics

Healthcare artificial intelligence: the road to hell is paved with good intentions

Usman Iqbal [1,2] Leo Anthony Celi [3,4,5] Yi-Hsin (Elsa) Hsu,[6,7] Yu-Chuan (Jack) Li [8,9,10]

To cite: Iqbal U, Celi LA, Hsu Y-HE, *et al*. Healthcare artificial intelligence: the road to hell is paved with good intentions. *BMJ Health Care Inform* 2022;**29**:e100650. doi:10.1136/bmjhci-2022-100650

Received 18 July 2022
Accepted 21 July 2022

For numbered affiliations see end of article.

Correspondence to
Prof. Dr Yu-Chuan (Jack) Li; jack@tmu.edu.tw

The *BMJ Health & Care Informatics* presented two editors' choice papers highlighting artificial intelligence (AI) and the challenges to properly evaluating AI-driven implementation tools associated with healthcare improvement at the system level.

The study from Kueper *et al*[1] focused on AI challenges in the primary care setting in Ontario, Canada. They provided lessons learnt and guidance for future opportunities to improve primary care using AI for resource management. The authors engaged multistakeholders in collaborative consultations. Nine priorities were identified that centred on system-level considerations, such as practice context, organisation and a performance domain devoted to health service delivery and quality of care. The paper highlighted concerns around equity and the digital divide, system capacity and culture, data accessibility and quality, legal and ethical considerations, user-centred design, patient-centredness, and appropriate assessment of AI application.

The role of AI within the learning health system framework is reviewed. AI models should be developed and applied to healthcare processes safely and meaningfully to optimise system performance and the society's well-being.[2] Moreover, AI provides preventive and pre-emptive medicine opportunities that are most valuable when they are prompt, accurate, personalised and acted upon expeditiously.[3]

Sikstrom *et al*[4] analysed a broad range of literature and investigated the bias and disparities that emerge from the application of AI in medicine. In this study, the authors proposed three pillars (transparency, impartiality and inclusion) for health equity and clinical algorithms. In addition, they proposed a multidimensional conceptual framework to evaluate AI fairness in healthcare. This framework is designed to ensure that decision support tools that provide predictions promote health equity.

A crucial problem facing AI research is data focused on specific regions and diseases that are then used to validate and train the algorithms, resulting in lack of generalisability over the global AI research landscape.[5 6] There is growing evidence that AI tools that perpetuate or even magnify inequities and disparities are often due to design and development misspecifications. Standards and classification system for AI-based healthcare technologies are required to facilitate research and evaluation to mitigate unintended harm and maximise patient and systems benefits.[7 8] All stakeholders need to be involved in validating the feasibility and effectiveness of AI.

The application of AI in medicine faces several challenges. It requires a development lifecycle framework that prioritises health equity and social justice.[9 10] Ultimately, AI systems must be continuously monitored to ensure that it does not contribute to outcome disparities across patient demographics.

Author affiliations
[1]Global Health and Health Security Department, College of Public Health, Taipei Medical University, Taipei, Taiwan
[2]HealthICT, Department of Health, Tasmania, Australia
[3]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[4]Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[5]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[6]BioTech EMBA Program, International PhD Program in Biotech and Healthcare Management, School of Healthcare Administration, College of Management, Taipei Medical University, Taipei, Taiwan
[7]School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[8]Graduate Institute of Biomedical Informatics, College of Medical Science & Technology, Taipei Medical University, Taipei, Taiwan
[9]Dermatology Department, Wan-Fang Hospital, Taipei, Taiwan

[10]International Association of Medical Informatics (IMIA), Geneva, Switzerland

Twitter Usman Iqbal @UsmanIqbal85, Leo Anthony Celi @MITCriticalData and Yu-Chuan (Jack) Li @jaak88

Contributors Initial conception and design: UI, LA C, Y-HH, Y-C L. Drafting the manuscript: UI, LAC, Y-HH, Y-CL. Critical revision of the paper: UI, LAC, Y-HH, Y-CL.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

ORCID iDs

Usman Iqbal http://orcid.org/0000-0002-0614-123X
Leo Anthony Celi http://orcid.org/0000-0001-6712-6626
Yu-Chuan (Jack) Li http://orcid.org/0000-0001-6497-4232

REFERENCES

1 Kueper JK, Terry A, Bahniwal R, *et al*. Connecting artificial intelligence and primary care challenges: findings from a multi stakeholder collaborative consultation. *BMJ Health Care Inform* 2022;29:e100493.
2 Sujan M, Pool R, Salmon P. Eight human factors and ergonomics principles for healthcare artificial intelligence. *BMJ Health Care Inform* 2022;29:e100516.
3 Iqbal U, Celi LA, Li Y-CJ. How can artificial intelligence make medicine more preemptive? *J Med Internet Res* 2020;22:e17211.
4 Sikstrom L, Maslej MM, Hui K, *et al*. Conceptualising fairness: three pillars for medical algorithms and health equity. *BMJ Health Care Inform* 2022;29:e100459.
5 Celi LA, Cellini J, Charpignon M-L, *et al*. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digit Health* 2022;1:e0000022.
6 Zhang J, Whebell S, Gallifant J, *et al*. An interactive dashboard to track themes, development maturity, and global equity in clinical artificial intelligence research. *Lancet Digit Health* 2022;4:e212–3.
7 Shelmerdine SC, Arthurs OJ, Denniston A, *et al*. Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare. *BMJ Health Care Inform* 2021;28:e100385.
8 Unsworth H, Wolfram V, Dillon B, *et al*. Building an evidence standards framework for artificial intelligence-enabled digital health technologies. *Lancet Digit Health* 2022;4:e216–7.
9 Feng J, Phillips RV, Malenica I, *et al*. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. *NPJ Digit Med* 2022;5:66.
10 Dankwa-Mullan I, Scheufele EL, Matheny ME, *et al*. A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle. *J Health Care Poor Underserved* 2021;32:300–17.

BMJ Health & Care Informatics

Effect of digital-enabled multidisciplinary therapy conferences on efficiency and quality of the decision making in prostate cancer care

Erik Ronmark,[1] Ralf Hoffmann,[2] Viktor Skokic,[1,3] Maud de Klerk-Starmans,[2] Fredrik Jaderling,[3] Pieter Vos,[2] Maudy C W Gayet,[2] Hans Hofstraat,[2] Marco Janssen,[2] Olof Akre,[1,3] Per Henrik Vincent [1,3]

To cite: Ronmark E, Hoffmann R, Skokic V, *et al*. Effect of digital-enabled multidisciplinary therapy conferences on efficiency and quality of the decision making in prostate cancer care. *BMJ Health Care Inform* 2022;**29**:e100588. doi:10.1136/bmjhci-2022-100588

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2022-100588).

Received 25 March 2022
Accepted 12 July 2022

[1]Department of Pelvic Cancer, Karolinska University Hospital, Stockholm, Sweden
[2]Philips Research, Eindhoven, The Netherlands
[3]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

Correspondence to
Dr Per Henrik Vincent;
per.vincent@regionstockholm.se

ABSTRACT

Objectives To investigate the impact on efficiency and quality of preprostatectomy multidisciplinary therapy conferences (MDT) at Karolinska University Hospital related to the use of a digital solution compared with standard of care. Further, to explore whether gains in MDT efficiency and quality impact oncological or functional patient outcomes.

Methods We conducted a prospective, observational study of preoperative prostate cancer MDT at Karolinska between February 2017 and March 2021, including 1329 patients. We compared efficiency and quality of the standard MDT and the MDT using the digital solution IntelliSpace Precision Medicine Multidisciplinary Team Orchestrator (ISPM) based on the previously used MDT-MODe approach. Clinical and patient-reported functional outcomes were derived from the medical records and the Swedish National Prostate Cancer Register.

Results While ISPM was used during the MDT meeting, the time spent per patient was reduced by 24% (p<0.001) and most of the MDT-MODe items were scored significantly higher. There was a reduction in pelvic lymph-node dissection procedures in the ISPM cohort (p=0.001) and an increased proportion of unilateral nerve-sparing procedures (p=0.005), while all other outcome-related measures were not significantly different between the two patient groups.

Discussion and conclusion To increase the value of the MDT, all data relevant for treatment decision need to be purposefully presented and compiled, which also enables secondary use of the data.

The use of a digital solution during preoperative MDTs for prostate cancer decision making at Karolinska University Hospital improved the efficiency and quality of this multidisciplinary team meeting without impacting patient outcomes.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Multidisciplinary therapy conferences are widely used in modern cancer care and patients discussed in a multidisciplinary therapy conference are more likely to receive appropriate staging and treatment. However, the multidisciplinary therapy conference is time consuming and rarely digitalised or adequately structured.

WHAT THIS STUDY ADDS

⇒ The use of a digital clinical decision support system during preoperative prostate cancer multidisciplinary therapy conferences improved the efficiency and quality of the meetings but was not associated with changes in oncological and functional outcomes after surgery.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Apart from allowing for more efficient use of clinical resources, digitalisation of multidisciplinary therapy conferences holds a promise to enable truly data-driven clinical workflows.

INTRODUCTION

The multidisciplinary therapy conference (MDT) has become a corner stone of cancer care. Patients who are discussed in an MDT, where a team of hospital staff gather to summarise relevant data and decide on treatment recommendations, are more likely to receive appropriate staging and treatment plans, but it is unclear whether this also results in improved patient outcomes.[1] Among parameters that may affect the value of an MDT, leadership, clarity of objectives, technical equipment for visualisation and electronic documentation, continuous audit of the process, access to complete case information and clarified roles of healthcare professionals have been identified as potentially vital prerequisites for a systematic MDT approach.[2 3]

The MDT often gathers a large number of health professionals, and, with more complex diagnostic and therapeutic options, the quality and efficiency of the decision-making process becomes increasingly important.

MDT conferences are rarely fully digitalised or adequately structured, which may affect the quality and efficiency of the decision-making process.[4] Data are not compiled and presented visually in a structured way and clinical parameters are presented verbally, which may lead to delays in the discussion when information needs to be repeated. Lack of continuous access to the clinical parameters during the MDT session may lead to information loss and hamper the multidisciplinary character of the MDT, thereby increasing the risk of non-optimal treatment decisions.[5 6] Moreover, if the consensus decisions are not captured in the electronic medical records (EMR) in real-time, this may lead to errors, misunderstandings and delay in data transfer to the EMR.

With the field of digital health evolving rapidly, solutions for MDTs have been developed and assessed. Structuring MDTs by use of such solutions has been shown to increase adherence to national guidelines and efficiency in several tumour forms.[7 8]

To increase the MDT efficiency without compromising quality of patient care, multiple quality-assessment tools and discussion checklists have been developed. Whether these tools also positively impact patient outcomes remains unknown.[9] In this study, we hypothesised that use of a digital, patient-centric, diagnosis-specific solution developed jointly by us (IntelliSpace Precision Medicine Multidisciplinary Team Orchestrator, further referred to as 'ISPM' throughout this text) during preprostatectomy MDTs at Karolinska University Hospital would improve the efficiency and quality of the MDT. The primary aim of the study was to investigate whether the use of the ISPM application saved meeting time and improved the quality of the decision process. The secondary aim was to assess whether the oncological and functional patient outcomes were affected by the implementation of ISPM.

METHODS

We have done a prospective observational cohort study comparing patient cohorts before and after the introduction of the clinical decision-support tool ISPM. The study was conducted between February 2017 and March 2021 at Karolinska University Hospital including patients discussed at preprostatectomy MDTs before undergoing robot-assisted radical prostatectomy.

Study setting

Hospital care in Sweden is entirely funded by taxes, and is therefore, as a rule, population based. Karolinska University Hospital is a Swedish tertiary referral hospital treating patients in all risk categories but with emphasis on high-risk patients referred from all regions of Sweden.

The weekly preprostatectomy MDT meeting is attended by 10–12 specialists in urology and radiology and aims to find a surgical strategy for an optimal balance between radical removal of the prostate cancer and postoperative functional outcomes.

Before we introduced ISPM, staff urologists took turns chairing the MDT, verbally reporting the clinical data from printed EMR excerpts, followed by a presentation of the MR images by a radiologist. The staff then discussed the optimal strategy for degree of nerve-sparing surgery, extent of sphincter sparing dissection in the apex, lymph-node dissection or not, degree of radicality in the bladder neck and the seminal vesicles. The concluded surgical treatment strategy plans were documented by the respective chair urologist in the EMR after the conference.

After the introduction of ISPM, all relevant clinical and radiological data were entered in the ISPM platform prior to the MDT meeting. In contrast to the baseline setting, clinical and radiological data were continuously visualised on the ISPM dashboard during the MDT meeting alongside the MR images until the surgery treatment plan had been captured in ISPM using the treatment plan documentation tool of the application (figure 1).

A baseline measurement in the standard MDT setting (before the use of ISPM) was carried out (February 2017–Septmber 2019), and, consecutively, data were collected while ISPM was in use (October 2019–March 2021). The efficiency and quality of the MDTs was compared by timing the discussion and using a modified version of the Metric of Decision-Making (MDT-MODe).[10] Nine items measuring quality were scored using a Likert scale (1, 3 and 5) with higher score indicating higher quality (for details of the modified version of the MDT-MODe used in this study, see online supplemental table 1). We grouped the MDT-MODe items into two main categories: MDT-MODe items relating to the availability and presentation of decision-relevant data, and MDT-MODe items related to the efficiency of MDT execution and team member interaction. Two observers, not participating in the therapy discussion, took turns assigning the MDT-MODe scores. An inter-rater variability analysis was conducted by letting the two observers assign scores to the same MDTs on three separate occasions to ensure agreement.

Software platform

The 'ISPM' software solution enables preparing, scheduling, visualisation, presentation and documentation of information and decisions taken in MDT case discussions. Using SQL queries, the system collects and transforms structured and unstructured data from the hospital data lake into a prostate data model and stores the result into an FHIR database following SNOMED-CT codes. In the study implementation, variables of interest but not available in the research copy of the Karolinska data lake were manually entered in ISPM.

Patient population

In all, 924 patients were discussed at MDTs in the period February 2017–September 2019, before the implementation of the ISPM software ('baseline' cohort), and 405 at conferences between October 2019 and March 2021 using ISPM ('ISPM' cohort). Only patients undergoing prostatectomy as primary treatment for prostate cancer

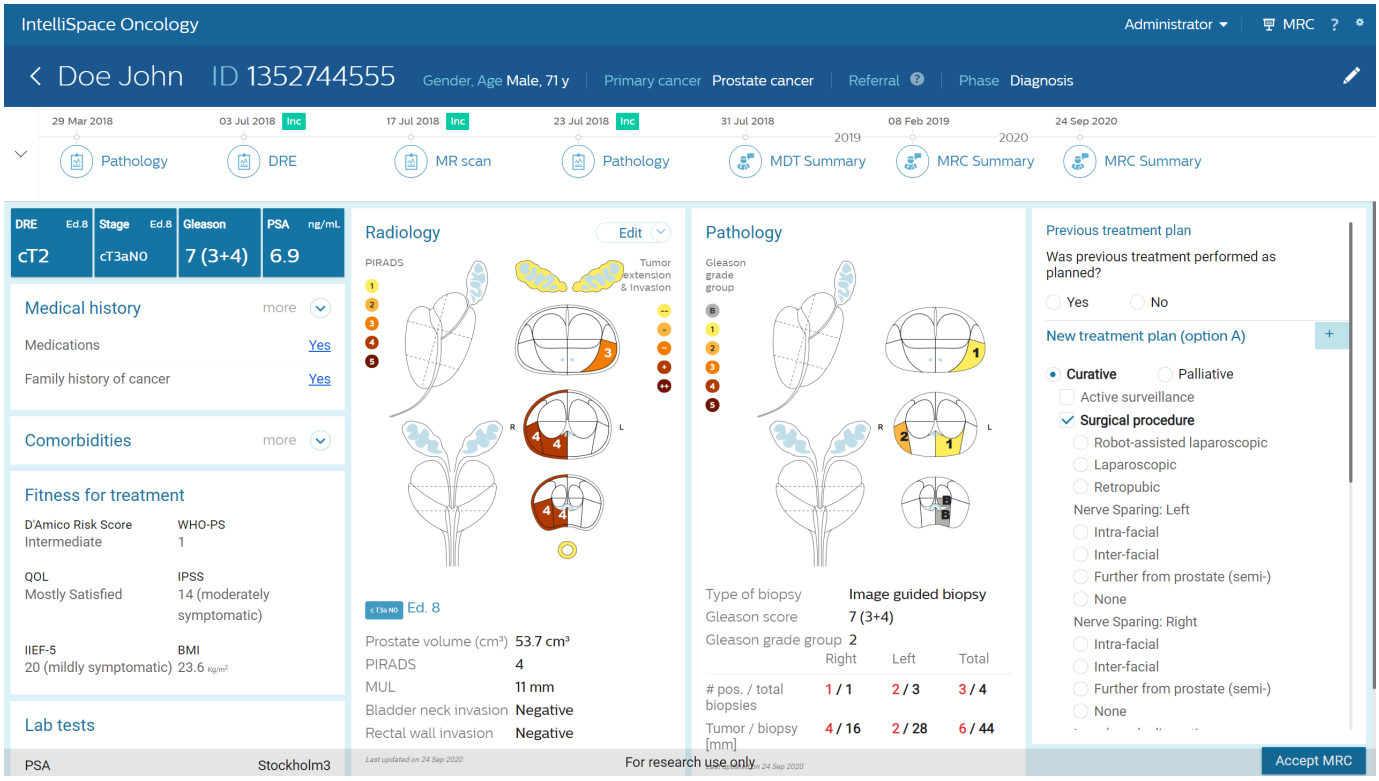

Figure 1 ISPM dashboard as implemented and used in the prospective, observational study on the impact of a digital solution during the MDTs in the prostate cancer care flow at Karolinska University Hospital. Patient data are fictional and do not originate from a real person. BMI, body mass index; IIEF-5, International Index of Erectile Function; MDT, multidisciplinary therapy; ISPM, IntelliSpace Precision Medicine Multidisciplinary Team Orchestrator; MDT, multidisciplinary therapy; QOL, quality of life; PSA, prostate-specific antigen; PIRADS, Prostate Imaging-Reporting & Data System; MRC, magnetic resonance imaging conference.

at Karolinska University Hospital within 30 days after their preoperative conference were included, to increase the likelihood that the conference decision was implemented. We assigned MDT-MODe scores to 164 baseline and 163 ISPM patients, at 21 and 22 MDTs, respectively.

Oncological and functional patient outcomes

All clinical and patient-reported outcome data were obtained from routinely collected clinical or quality follow-up data. Positive surgical margin (as sign of remaining cancer and hence non-radical treatment) was used as a surrogate for oncological quality with significant positive margin defined as at least of three millimetres length. Other relevant post-surgical and perisurgical outcomes, such as extended lymph-node dissection, positive lymph nodes, and nerve-sparing surgery, were also analysed. Functional outcomes were obtained using the questionnaires in the Swedish National Prostate Cancer Register (NPCR) that all prostate cancer patients in Sweden are invited to answer before undergoing primary treatment and twelve months after treatment. The questionnaires are administered in collaboration with the Swedish Regional Cancer Centres and NPCR and can be found at https://npcr.se/eprom/dokument. In this study, we defined urinary continence as 'use of less than one protective urinary pad per day' and urinary incontinence as 'use of one or more protective urinary pad per day'. Erectile function was measured using the International Index of Erectile Function questionnaire (IIEF-5)[11] with erectile dysfunction defined as less than 12 points. Quality of life regarding 'erectile function satisfaction and continence satisfaction' was defined as a self-report of either not bothering the patient at all or only to a small degree. Tumour grade was scored using ISUP grading.[12]

Statistical analysis

In tables 1 and 2 (and online supplemental table 2,3), comparisons of the characteristics of the studied population with respect to the use of ISPM at MDTs were structured according to the following: the distributions of numerical variables or ordinal variables with more than two levels were compared using the Mann-Whitney U test. The distributions of categorical variables with more than two categories were compared using the χ^2 test, whereas the distributions of categorical variables with two categories where one category was identified as the outcome of interest were compared in terms of prevalence ratios and the likelihood ratio test associated with an estimated log-binomial model. Levene's test, centred at the median, was used to assess the difference in variance between non-normally distributed variables.

For figures 2A and 3A–3D, the distributions of ordinal variables were compared using the Mann-Whitney U test. For figure 2B, the association between the usage of ISPM and the duration of discussion for each individual patient at the MDTs was studied using a linear regression model including the number of patients evaluated at a conference, the usage of ISPM and their interaction as explanatory variables.

Table 1 Patient demographics of the baseline versus the ISPM cohort

	Baseline	ISPM	P value
No of patients	n=924 (69.5%)	n=405 (30.5%)	
No of patients in MDT-MODe	n=164 (50.2%)	n=163 (49.8%)	
No of patients per conference	Mean=7.8 (SD=2.9)	Mean=7.4 (SD=2.6)	M-W p=0.74
No of staff per conference	Mean=11.7 (SD=2.7)	Mean=11.5 (SD=2.8)	M-W p=0.85
Patient age (years)	Mean=65.5 (SD=7.4)	Mean=65.9 (SD=7.1)	M-W p=0.48
Postoperative ISUP grade group			M-W p=0.27
ISUP 1	74 (8.3%)	11 (3.1%)	
ISUP 2	424 (47.6%)	181 (51.3%)	
ISUP 3	274 (30.8%)	118 (33.4%)	
ISUP 4	44 (4.9%)	15 (4.2%)	
ISUP 5	74 (8.3%)	28 (7.9%)	
Missing*	34 (3.7%)	52 (12.8%)	
Postoperative T stage (pT)			M-W p=0.34
pT2	535 (59.8%)	215 (61.3%)	
pT3a	258 (28.8%)	111 (31.6%)	
pT3b	100 (11.2%)	25 (7.1%)	
pT4	2 (0.2%)	0 (0%)	
Missing*	29 (3.1%)	54 (13.3%)	
Preoperative incontinence			M-W p=0.008
Continent	296 (96.7%)	289 (99.7%)	
Incontinent	10 (3.3%)	1 (0.3%)	
Missing*	618 (66.9%)	115 (28.4%)	
Preoperative IIEF-5 score			M-W p=0.90
	Mean=14.2 (SD=9.5) (#missing*=618 (66.9%))	Mean=14.4 (SD=9.6) (#missing*=122 (30.1%))	
Preoperative erectile dysfunction - outcome: impotence (IIEF-5 score <12)			LRT p=0.86
Frequencies	133/306 (43.5%) (#missing*=618 (66.9%))	121/283 (42.8%) (#missing*=122 (30.1%))	
Prevalence ratios	1.0 (Ref.)	0.98 (0.81–1.18)	
Response frequency to preoperative questionnaire			χ^2 p<0.00 1
	309 (33.4%)	290 (71.6%)	

*Percentage missing calculated on the entire cohort.
IIEF-5, International Index of Erectile Function questionnaire; ISPM, IntelliSpace Precision Medicine Multidisciplinary Team Orchestrator; ISUP, International Society of Urological Pathology; LRT, likelihood ratio test; MDT, multidisciplinary therapy conference; M-W, Mann-Whitney U test.

All calculations were performed in R V.4.0.0.

We have used the SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence) guidelines in preparation for this manuscript.[13]

RESULTS

Patient cohorts

The two cohorts (baseline and ISPM) were similar with respect to demographic and clinical characteristics, mean patient age, postoperative ISUP grade group, tumour stage, and erectile function (table 1). The response rate to the NPCR questionnaires measuring preoperative functional status was significantly higher among patients discussed using the ISPM solution, 70.9% vs 32.6% when ISPM was not used (p<0.001). Among those who responded, incontinence was more common in the baseline group, but the difference was small in absolute numbers, 3.4% (10 out of 298 patients) compared with 0.3% (1 out of 287 patients).

Time efficiency of the MDT

The average time spent discussing each patient was 24% shorter in the ISPM compared with the baseline setting (3.8 vs 5.0 min; p<0.001; figure 2A). There was also a significant difference in variances between comparison

Table 2 Oncological, perioperative and 12-month functional (urinary, sexual) patient outcomes of the baseline versus the ISPM patient cohort

	Baseline	ISPM	P value
Positive surgical margin			LRT 0.51
Frequencies	251/906 (27.7%) (#missing=18 (1.9%))	98/378 (25.9%) (#missing=27 (6.7%))	
Prevalence ratios	1.0 (Ref.)	0.94 (0.76–1.14)	
Nerve-sparing—any			M-W=0.11; $\chi 2$=0.13
Any nerve-sparing	805 (88.2%)	353 (91.2%)	
No nerve-sparing	108 (11.8%)	34 (8.8%)	
Missing*	11 (1.2%)	18 (4.4%)	
Nerve-sparing unilaterally or bilaterally—dichotomised, with pairwise comparisons			c^2=0.005 – OVERALL
Bilateral	392 (42.9%)	140 (36.2%)	c^2=0.05†
Unilateral	413 (45.2%)	213 (55.0%)	c^2=0.005†
No nerve-sparing	108 (11.8%)	34 (8.8%)	c^2=0.13†
Missing*	11 (1.2%)	18 (4.4%)	
Pelvic lymph-node dissection			M-W p<0.001
No	543 (60.0%)	248 (70.1%)	
Yes	362 (40.0%)	106 (29.9%)	
Missing*	19 (2.1%)	51 (12.6%)	
Lymph-node metastases among patients that underwent pelvic lymph-node dissection			M-W p=0.92
No (N0)	302 (83.4%)	88 (83.0%)	
Yes (N1)	60 (16.6%)	18 (17.0%)	
Missing*	0 (0%)	0 (0)	
Erectile dysfunction at 12 months after surgery—outcome: impotence (IIEF-5 score <12)			LRT 0.90
Frequencies	451/604 (74.7%) (#missing*=320 (34.6%))	172/229 (75.1%) (#missing*=176 (43.5%))	
Prevalence ratios	1.0 (Ref.)	1.01 (0.92–1.09)	
Incontinence at 12 months after surgery—outcome: incontinence			LRT 0.98
Frequencies	171/619 (27.6%) (#missing*=305 (33.0 %))	64/231 (27.7%) (#missing*=174 (43.0%))	
Prevalence ratios	1.0 (Ref.)	1.0 (0.78–1.27)	

*Percentage missing calculated on the entire cohort.
†Pairwise comparison; Bonferroni-Holm corrected for multiple testing.
IIEF-5, International Index of Erectile Function; ISPM, IntelliSpace Precision Medicine Multidisciplinary Team Orchestrator; LRT, likelihood ratio test; M-W, Mann-Whitney U test.

groups (Levene's test; p<0.001), indicating a more predictive duration per patient when using ISPM. During the baseline period, the time spent discussing each patient decreased with increasing number of cases in the meeting (Pearson correlation=−0.23; p=0.04). During the ISPM period, there was no such significant correlation (Pearson correlation=−0.075; p=0.36; figure 2B).

Quality of the MDT

There were higher MDT-MODe scores for the information presentation items psychosocial, comorbidity (p<0.01) and pathology (p<0.05) during the ISPM period, while there was no difference in the presentation of patient's views, imaging, and patient history (figure 3A). Team interaction items regarding quality of leadership (Chair) and contribution of specialty (Members) also received higher scores in the ISPM setting (p<0.001; figure 3B). Furthermore, the fraction of participants actively taking part in the MDT discussion increased using ISPM (p<0.05; figure 3C). Moreover, we observed that there were significantly fewer questions on already presented data raised in the meeting while ISPM was in use (p<0.01; figure 3D).

Oncological and functional outcomes

There was no statistically significant difference with respect to oncological outcomes between the baseline and the ISPM cohorts. The proportion of men with positive surgical margins was 27.7% in the baseline group and 25.9% in the ISPM group (p=0.66; table 2; online supplemental table 2).

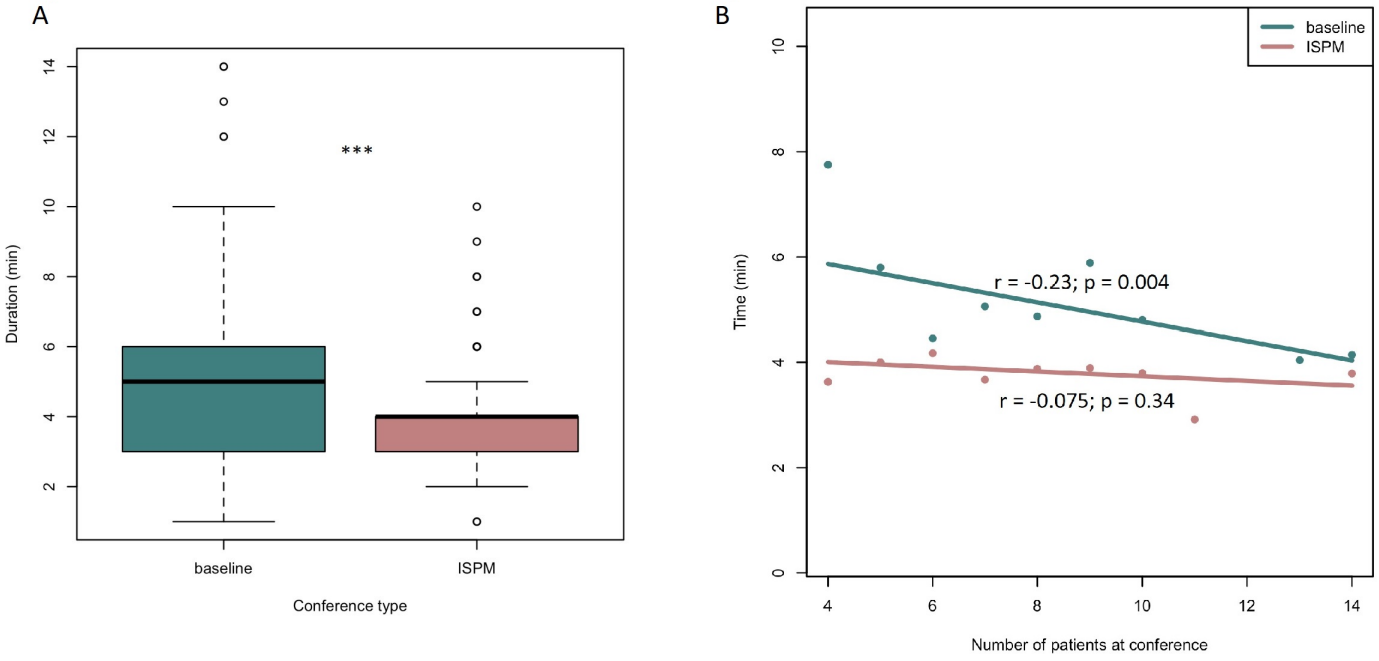

Figure 2 A Time spent in the MDT meeting per patient in the baseline setting (164 patients) versus the ISPM setting (163 patients). Box plot with median and IQR; whiskers denote $\pm 1.5 \times$IQR. ***P<0.001 (B)**:** interaction between mean time (minutes) spent per patient and number of patients scheduled and discussed during the MDT. Dots indicate the mean durations at conferences with a particular number of patients being discussed. Regression lines are derived from 164 (baseline) and 163 (ISPM) patients per group. ISPM, IntelliSpace Precision Medicine Multidisciplinary Team Orchestrator; MDT, multidisciplinary therapy conference.

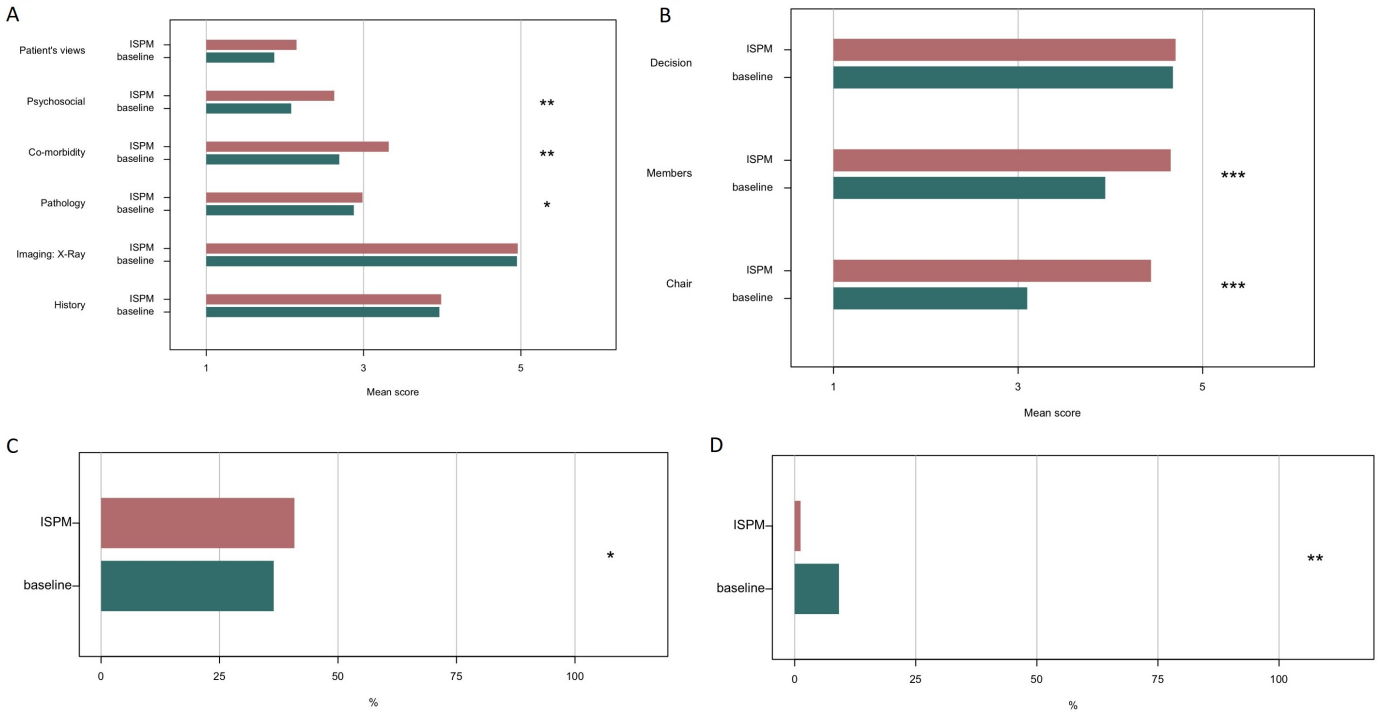

Figure 3 MDT-MODe items concerning information presentation. *P<0.05, **p<0.01. (B) MDT-MODe items concerning leadership and team interaction. ***P<0.001. (C) Percentage of staff members actively participating per patient case discussion and decision making in the MDT. *P<0.05 (D) Percentage of patients for which questions were raised during the MDT meeting to repeat already presented information. **P<0.01. ISPM, IntelliSpace Precision Medicine Multidisciplinary Team Orchestrator; MDT, multidisciplinary therapy conference; MDT-MODe, Metric of Decision-Making.

The overall frequency of nerve-sparing and non-nerve-sparing surgery was virtually similar in the comparison groups, but the patterns differed slightly with more unilateral nerve-sparing surgery in the ISPM group and more bilateral nerve-sparing and non-nerve-sparing in the baseline group (table 2).

An extended pelvic lymph-node dissection was carried out more often in the baseline cohort, 39.1% vs 26.4% for the ISPM cohort, respectively (p=0.001; table 2), but there was no difference in proportion of histologically confirmed metastases (16.6% in the baseline vs 17.0% in the ISPM cohort; p=0.92; table 2).

The functional outcomes 1 year after surgery were similar for the two groups. Erectile dysfunction (IIEF-5 score <12) at 12 months was present among 74.4% (baseline) and 75.1% (ISPM) (p=0.90; table 2) of patients, and incontinence (daily use of one or more urinary pads) was present among 27.6% (baseline) and 27.7% (ISPM) (p=0.98; table 2; online supplemental table 3).

DISCUSSION

We found that implementing ISPM in a multidisciplinary tumour conference was associated with increased efficiency of the conference and less time needed to discuss each patient case. The results also indicated that the variability of the length of each case discussion was reduced with the use of ISPM. At the same time, the quality of the teamwork and the decision-making process was improved with the use of ISPM. These improvements were not, however, reflected in improved oncological or functional patient outcomes.

Using ISPM, discussing one patient took on average 72 s less, which corresponds to a 24% reduction, or approximately 9–10 min shorter MDT in the current setting. Considering that there was a mean of 11.6 participants, more than one person-hour was saved during each session. This time saving is in agreement with results from another group developing a similar oncological clinical decision support system for other cancer types.[8 14]

MDTs are in general scheduled events with a finite duration while the number of patient cases fluctuates and a structured process for the presentation and discussion of each case is paramount for retaining the quality of the decision making throughout the conference. We found both decreased variability of the duration of each case discussion and a consistent duration per case regardless of the number of patients discussed at the conference, indicating that the use of ISPM leads to a more structured and predictable process.

Although the quality of the MDTs increased when ISPM was used, this was not reflected in improved oncological or functional patient outcomes in our data. It should be noted that, already prior to the implementation of ISPM, the format of the preoperative prostate-cancer MDTs at Karolinska had been structuralised—although not into a digital format—with an apparent effect on the nerve-sparing strategy as well as on the risk of positive surgical margins.[15]

While it has been shown that MDTs lead to more accurate staging,[16] higher adherence to clinical guidelines,[17] and shorter time to treatment after diagnosis,[18] several prior studies have failed to show improved outcome among patients discussed in MDT meetings[19–21] while other have reported better outcomes.[22–24] The MDT is a costly process, and it is important that future studies justify the costs through evidence of better outcomes.

A structured digital format for the MDT entails several potential further advantages apart from efficiency and quality in the decision making. For example, the resulting database can be used for real-time quality assessment, feedback to pathologists, radiologists, surgeons and radiotherapists. Also, it enables development of prognostic models for better prediction tailored to the centres' own patient cohorts. Patient-reported outcome measures can be used in the communication with the patient during follow-up for a more structured care of the side effects of treatment and for spending more time with the patient on solving problems rather than understanding them. None of these advantages were assessed in this study but are all strong potential benefits of a digital platform such as ISPM.

Without simultaneous evaluation of positive surgical margins and functional outcomes, quality assessment of prostate cancer surgery is of little use since there is a reciprocal relation between radicality and postoperative function. Digital platforms connecting data points on all dimensions will facilitate more precise quality assessment. Ultimately, applying deep learning to make fuller use of these rich clinical, morphological and patient-reported data is a promising future development.

The main limitation of this study is the observational design with non-concurrent comparison groups. The baseline measurement was carried out over a period of 33 months before the ISPM solution was implemented. Both treatment and outcome of prostate cancer change over time[25] and differences between the baseline and ISPM periods may be attributable to other time-varying factors, such as staff turnover. Furthermore, the lower frequency of pelvic lymph-node dissections in the ISPM period may, apart from a true effect of using the digital platform, be due to subtle changes in our operative indications for the procedure.

Access to patient-reported data is a major clinical need in healthcare in general, but particularly in the care of prostate cancer. The response rate to preoperative questionnaires was low in the baseline group, 32.6% compared with 70.9% the ISPM group. This difference reflects our effort made during the study period to increase patient participation in the national questionnaires on functional outcome rather than an effect of the digital platform.

Conclusion

Our implementation of the ISPM clinical decision support system in MDT sessions at Karolinska University Hospital was associated with more efficient presentations

and decision making in the conference as well as higher perceived quality of the decision process, but not with improved patient outcomes.

Contributors RH, MdK-S, FJ, PV, HH, MJ, OA and PHV conceived the study; RH, PV, OA and PHV wrote the study plan with input from ER, FJ, HH, MJ; ER, RH, MdK-S, FJ, PV, HH, MJ, OA and PHV contributed to the design of the software; ER, VS, FJ, PV, MdK-S, PHV collected the data; ER, VS, PHV performed the statistical analyses with input from RH, MCWG; ER, MCWG drafted the manuscript; all coauthors contributed substantially to the revision of the manuscript and approved the final version; ER and RH contributed equally to this paper; OA is principal investigator of the clinical study and guarantor of the work.

Funding This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780495.

Competing interests RH, MdK-S, PV, MCWG, HH and MJ are employees of Philips Research, Eindhoven, Netherlands.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the Swedish Ethical Review Authority (Dnr 2018/109-31; 2020-05939).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

ORCID iD

Per Henrik Vincent http://orcid.org/0000-0002-9895-5803

REFERENCES

1 Pillay B, Wootten AC, Crowe H, *et al*. The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: a systematic review of the literature. *Cancer Treat Rev* 2016;42:56–72.
2 Raine R, Wallace I, Nic a' Bhaird C. Improving the effectiveness of multidisciplinary team meetings for patients with chronic diseases: a prospective observational study. *Southampton* 2014.
3 Rosell L, Alexandersson N, Hagberg O, *et al*. Benefits, barriers and opinions on multidisciplinary team meetings: a survey in Swedish cancer care. *BMC Health Serv Res* 2018;18:249.
4 Janssen A, Robinson T, Brunner M, *et al*. Multidisciplinary teams and ICT: a qualitative study exploring the use of technology and its impact on multidisciplinary team meetings. *BMC Health Serv Res* 2018;18:444.
5 Horlait M, Baes S, Dhaene S, *et al*. How multidisciplinary are multidisciplinary team meetings in cancer care? an observational study in oncology departments in Flanders, Belgium. *J Multidiscip Healthc* 2019;12:159–67.
6 Soukup T, Murtagh G, Lamb B, *et al*. Degrees of Multidisciplinarity underpinning care planning for patients with cancer in Weekly multidisciplinary team meetings: conversation analysis. *J Multidiscip Healthc* 2021;14:411–24.
7 Farrugia DJ, Fischer TD, Delitto D, *et al*. Improved breast cancer care quality metrics after implementation of a standardized tumor board documentation template. *J Oncol Pract* 2015;11:421–3.
8 Hammer RD, Fowler D, Sheets LR, *et al*. A digital tumor board solution impacts case discussion time and postponement of cases in tumor boards. *Health Technol* 2021;11:525–33.
9 Brown GTF, Bekker HL, Young AL. Quality and efficacy of multidisciplinary team (MDT) quality assessment tools and discussion checklists: a systematic review.. *BMC Cancer,* 2021.
10 Lamb BW, Wong HWL, Vincent C, *et al*. Teamwork and team performance in multidisciplinary cancer teams: development and evaluation of an observational assessment tool. *BMJ Qual Saf* 2011;20:849–56.
11 Rosen RC, Cappelleri JC, Smith MD, *et al*. Development and evaluation of an abridged, 5-item version of the International index of erectile function (IIEF-5) as a diagnostic tool for erectile dysfunction. *Int J Impot Res* 1999;11:319–26.
12 Epstein JI, Egevad L, Amin MB, *et al*. The 2014 International Society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. *Am J Surg Pathol* 2016;40:244–52.
13 Ogrinc G, Mooney SE, Estrada C, *et al*. The Squire (standards for quality improvement reporting excellence) guidelines for quality improvement reporting: explanation and elaboration. *Qual Saf Health Care* 2008;17 Suppl 1:i13–32.
14 Krupinski EA, Comas M, Gallego LG, *et al*. A new software platform to improve multidisciplinary tumor board Workflows and user satisfaction: a pilot study. *J Pathol Inform* 2018;9:26.
15 Jäderling F, Akre O, Aly M, *et al*. Preoperative staging using magnetic resonance imaging and risk of positive surgical margins after prostate-cancer surgery. *Prostate Cancer Prostatic Dis* 2019;22:391–8.
16 Sundi D, Cohen JE, Cole AP, *et al*. Establishment of a new prostate cancer multidisciplinary clinic: format and initial experience. *Prostate* 2015;75:191–9.
17 Korman H, Lanni T, Shah C, *et al*. Impact of a prostate multidisciplinary clinic program on patient treatment decisions and on adherence to NCCN guidelines: the William Beaumont Hospital experience. *Am J Clin Oncol* 2013;36:121–5.
18 Yopp AC, Mansour JC, Beg MS, *et al*. Establishment of a multidisciplinary hepatocellular carcinoma clinic is associated with improved clinical outcome. *Ann Surg Oncol* 2014;21:1287–95.
19 Boxer MM, Vinod SK, Shafiq J, *et al*. Do multidisciplinary team meetings make a difference in the management of lung cancer? *Cancer* 2011;117:5112–20.
20 Swellengrebel HAM, Peters EG, Cats A, *et al*. Multidisciplinary discussion and management of rectal cancer: a population-based study. *World J Surg* 2011;35:2125–33.
21 Wille-Jørgensen P, Sparre P, Glenthøj A, *et al*. Result of the implementation of multidisciplinary teams in rectal cancer. *Colorectal Dis* 2013;15:410–3.
22 Kočo L, Weekenstroo HHA, Lambregts DMJ, *et al*. The effects of multidisciplinary team meetings on clinical practice for colorectal, lung, prostate and breast cancer: a systematic review. *Cancers* 2021;13. doi:10.3390/cancers13164159. [Epub ahead of print: 18 08 2021].
23 MacDermid E, Hooton G, MacDonald M, *et al*. Improving patient survival with the colorectal cancer multi-disciplinary team. *Colorectal Dis* 2009;11:291–5.
24 Palmer G, Martling A, Cedermark B, *et al*. Preoperative tumour staging with multidisciplinary team assessment improves the outcome in locally advanced primary rectal cancer. *Colorectal Dis* 2011;13:1361–9.
25 Orrason AW, Westerberg M, Garmo H, *et al*. Changes in treatment and mortality in men with locally advanced prostate cancer between 2000 and 2016: a nationwide, population-based study in Sweden. *BJU Int* 2020;126:142–51.