COllaborative open platform E-cohorts for research acceleration in trials and epidemiology

Viet-Thi Tran, Philippe Ravaud

▶ **To cite this version:**

Viet-Thi Tran, Philippe Ravaud. COllaborative open platform E-cohorts for research acceleration in trials and epidemiology. Journal of Clinical Epidemiology, 2020, 124, pp.139-148. 10.1016/j.jclinepi.2020.04.021 . hal-03346351

HAL Id: hal-03346351
https://hal.inrae.fr/hal-03346351v1

Submitted on 6 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S089543561930976X
Manuscript_a73a67db57d1f422a6877a34d2ac861c

COllaborative Open Platform E-cohorts for Research Acceleration in Trials and Epidemiology (COOPERATE)

Viet-Thi Tran[1, 2], Philippe Ravaud[1, 2, 3]

[1] Université de Paris, CRESS, INSERM, INRA, F-75004 Paris, France
[2] Center for Clinical Epidemiology, Hôtel-Dieu Hospital, Assistance Publique-Hôpitaux de Paris (AP-HP), 1 Place du Parvis Notre Dame, 75004 Paris, France
[3] Department of Epidemiology, Columbia University Mailman School of Public Health, 22 W 168th St, New York, NY, USA

Correspondence to:

Viet-Thi Tran - MD, PhD; Philippe Ravaud – MD, PhD
Hôpital Hôtel Dieu, Centre d'Épidémiologie Clinique, Paris, France
1 place du Parvis Notre-Dame, Paris 75181, France
Tel: +33 1 42 34 89 87, fax: +33 1 42 34 87 90, email: thi.tran-viet@aphp.fr

Words: 3308

Key words:

Epidemiologic Research Design

Cohort Studies

© 2020 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

Abstract (250/250)

Background

The current clinical research system relies on a "one-off" project-by project model involving a costly and time-wasting permanent construction and deconstruction of the research infrastructure. We propose a new model of research relying on collaborative principles: The COllaborative Open Platform (COOP') e-cohort.

Development

The COOP' e-cohort aims at building a large community of patients willing to participate in research by contributing to the generation of a large database of patient-reported data, passively enriched, at the individual level, by linkage with routinely collected care and/or medico-administrative data. Approved teams can use the platform and benefit from already enrolled participants or collected data or add new online questionnaires to perform observational or interventional studies to answer a broad range of research questions.

Application

The Community of Patients for Research (ComPaRe) is a proof-of-concept COOP' e-cohort in the field of chronic conditions that was launched in 2017. As of April 2020, 36 000 patients have joined the project and contributed to more than 4 million data points. Patient-reported data will be enriched by linkage with the French national health system databases and with hospital data for patients receiving care in the Paris region. Since 2017, 150 researchers have used the platform for research projects. Three clinical trials nested in ComPaRe have been funded.

Conclusion

By moving from myriad independent studies to a large collaborative infrastructure of research, COOP' e-cohorts will accelerate the research process by avoiding the redundancy of many steps common to all research projects and by limiting waste of research.

Today's clinical research system is in crisis. An evolution toward a more efficient and sustainable model of research is necessary. In this paper, we discuss the limits of the current clinical research system and propose a new design to accelerate research in chronic conditions: the COllaborative Open Platform E-cohort for Research Acceleration in Trials and Epidemiology (COOPERATE, or COOP' e-cohort).

Limits of the current clinical research model

The current system relies on an inefficient "one-off" project model

The current clinical research system relies on a "one-off" project model [1]. Prospective studies, whether they are observational or interventional, start from scratch and require years to achieve full enrollment. After the publication of results, they are put away, resulting in a constant construction and deconstruction of the clinical research infrastructure [2]. Avoiding this "one-off" project model could spare the never-ending collection of the same data (e.g., a clinical trial collects an average 169 case report form pages [3, 4]; with up to 80% of items collected not reported in final publications and already available from previous research [5-8]); reduce the burden of research for participants [9], and reduce the costs of research (e.g., a prospective cohort of 659 patients with myelodysplastic syndromes costs €22.05 M (€1.23 M per year) [10, 11]).

The evidence is insufficiently robust

Most research studies are underpowered because they fail to recruit their original target within the time originally specified [12]. This situation is not due to the unwillingness of patients to participate in research but rather to the lack of collaboration between researchers. For example, in rare diseases, 40% of prospective research studies involve a single centre with a limited number of participants (42 to 282 participants) [13]. Beyond sample size, current data sources currently used in research may also have flaws. For example, a large proportion of results from recent observational studies relying on data collected for care or reimbursement purposes may

be questioned because important missing covariates were not collected with careful measurement as part of a research protocol [14].

The evidence does not reflect the "real world" and its relevance for patients is limited

In many clinical research studies, participants are often different from "real-world" people because they are mainly those who obtain care in large hospitals which participate in research [15]. Further, some inclusion/exclusion criteria used may limit the validity of findings from clinical research. For example, in clinical trials, older, obese or multimorbid patients are frequently excluded even though they represent an important proportion of patients for whom research conclusions will be applied [16-18].

Patients are also rarely involved in the elaboration and conduct of research, thus creating a gap between researchers' and patients' priorities. For osteoarthritis of the knee, 60% of trials test medications, whereas only 11% of patients consider that research on medications is of high priority [19]. Outcomes chosen in clinical research may not be those considered important by patients. For example, in diabetes, only 18% of clinical trials used outcomes considered important by patients [20].

The current system is built on a tournament model

Issues of the current clinical research model also come from researchers' mindsets and habits. Especially, poor collaboration between people and institutions is responsible for redundant and fragmented research, small sample sizes and siloed research. The system is organized in silos by disease, specialty and/or methods used; with, for example, a distinct separation between qualitative and quantitative research and between interventional and observational designs. Despite actions from institutions and scientific journals, principles of data sharing and open research are rarely implemented in practice. Among the reasons, we may cite the disproportionate requirements from regulatory and governance bodies and/or the extensive

efforts required by investigators (e.g., development of consent procedures that permit data sharing, de-identification of study data, creation of detailed code books, and support for end data users) to allow data sharing [21, 22]. However, beyond these reasons, we may argue that the current model of research is tainted by greed: investigators "play in a tournament" in which they compete for funding in a world in which small differences in innovation can be magnified into large differences in recognition, financial reward, and career advancement [22].

A new model for the clinical research infrastructure

To address these problems, we propose a new model for the clinical research infrastructure: the Collaborative Open Platform (COOP') e-cohort. In this paper, we present the general principles of COOP' e-cohorts and illustrate how they may be implemented in practice with examples from our proof-of-concept project: the Community of Patients for Research (ComPaRe) (**Box 1 & 2**).

Building a community of patients for research

Patient-powered research networks such as PatientsLikeMe have demonstrated that patients' motivation and engagement could accelerate research, minimize recruitment issues, and improve the relevance of findings [23, 24]. The concept of the COOP' e-cohort is to build a large community of patients willing to participate in research about their chronic conditions by contributing to the generation of a large database of patient-reported data that is passively enriched by other sources of data routinely collected during care or for collected medico-administrative purposes.

In this community, patients go beyond "simple participation" and answers to online questionnaires. They become active members of the research community and are involved in every step of the research such as the governance of the program, the choice of projects to be conducted within the COOP e-cohort, the conception of all materials destined to participants,

data analysis, or the dissemination of the project. Involvement of patients in research allows results to be more relevant and practicable, and ultimately, to be used by society [25]. In ComPaRe, we systematically ask participants about their expectations for participating in a large participative research project after their first year of participation. In 2018, among the 6089 participants who joined the cohort, 2692 (44%) wanted to participate beyond answering questionnaires. To meet these needs, we set up multiple interactive information technology interfaces that allowed patients to actively participate in the construction of the project. One example is to propose research ideas by using the online platform INSPIRE (https://inspire-compare.fr/). INSPIRE uses questionnaires with open-ended questions with which participants can propose research ideas, in their own words, without any health/research literacy requirements. Open-text data are analyzed by using natural language processing methods to generate a list of "research ideas" organized by themes and diseases that are accessible to all participants and researchers in ComPaRe. We also designed testing environments in which volunteer patients could review and propose modifications to all materials (information, questionnaires, etc.) destined for other patients.

Developing a large-scale recruitment strategy

To generate "real-world" data at scale, COOP' e-cohorts should rely on far-reaching recruitment strategies with the following:

- Minimal eligibility and exclusion criteria to generate evidence in a population of patients close to those who will benefit from the research, especially multi-morbid patients.
- Direct outreach to potential participants by using widespread advertising in general and social media and partner patient associations. This strategy avoids the potential pitfalls of relying on clinicians' referral for recruitment (e.g., their lack of time and resources to cope with the additional workload of research duties) [26, 27]. Direct outreach

strategies have been used in initiatives such as All of Us [26], the UK Biobank [28], and the Nutrinet-Santé study [29].

- Scalable recruitment strategies. "Low-cost" communication channels enabling outreach to a large number of potential participants should be favored. For example, in ComPaRe, we systematically add information about the research program to e-mails sent to patients who booked an appointment online in one of our partner hospitals, which allowed for the steady recruitment of 500 participants with diverse chronic conditions every month. In addition, because most patients with one chronic condition have (or will develop) multiple chronic conditions, all efforts directed to the recruitment of patients with a given condition will also contribute to the recruitment of patients with other conditions.

Sharing a common and permanent infrastructure for simpler and faster research

The core concept of the COOP' e-cohort is to result in a collaborative and permanent infrastructure for research.

- This single and shared research infrastructure, involving a mutualized participant and data pool, facilitates and accelerates research for all. In a COOP' e-cohort, approved researchers can access the platform, use any data already collected or actively collect new data (e.g., by the addition of new online questionnaires) to conduct a wide range of studies. A variety of designs are possible, ranging from observational studies [30]; online qualitative studies [31]; collective intelligence studies [32]; development and validation of new patient-reported outcome measures (PROMs) [33], algorithms or scores; or even trials testing online tools, apps or educative programs, using cohort multiple randomized controlled trial designs [34]) (**Table**). Researchers using the COOP' e-cohorts benefit by 1) avoiding the numerous procedural and regulatory steps necessary to initiate a new project that have already been completed at the COOP' e-cohort level; 2) participants already recruited; and 3) data already collected (**Figure 1**).

- The single and shared large infrastructure allows for reaching a critical size that opens possibilities not conceivable with smaller one-off projects such as obtaining regulatory authorizations for linkage with administrative and/or hospital databases.
- Shared participant pool. Participants are recruited once and contribute to multiple research projects. For example, a participant joining the cohort to accelerate research on asthma will also contribute to accelerating research on the participant's other conditions.
- Shared data pool. Data are collected once but are used for multiple purposes. All data collected by researchers are made available to others. The more nested projects, the richer the data and the fewer additional specific variables that will be needed for future projects. The common and shared infrastructure also solves most "technical" barriers for data sharing. No time and effort is required from researchers to document the data (documentation is automatically generated by the platform for all patient reported data) [35].
- The shared infrastructure of the COOP' e-cohort thus reduces some costs related to study size. Costs are related mainly to the maintenance of the infrastructure rather than "per-study" costs [36].

Multiple data sources

COOP' e-cohorts combine active data collection by using online questionnaires and the passive enrichment of these patient-reported data with data from other sources (**Figure 2**).

Active data collection

Active data collection involves mainly data reported by participants themselves with online self-reported questionnaires to capture PROMs and patient-reported experience measures (PREMs). Electronic data collection allows for a naturalistic and flexible follow-up of patients who can answer questionnaires where and when they choose, rather than "at the next visit".

Data collected by online questionnaires may cover the following: 1) sociodemographic and clinical data (e.g., precise list of chronic conditions with diagnosis dates, all pharmacological and non-pharmacological treatments taken by patients); 2) PROMs or PREMs (e.g., burden of treatment, patient-reported adherence to medications); 3) stated preferences collected by using discrete choice experiments or case vignette tools; or 4) unstructured textual data collected by questionnaires with open-ended questions used to explore participants' perspectives on specific topics or engage them in the generation of new insights.

Self-reported questionnaires may target all participants from the COOP' e-cohort or sub-groups (or sub-cohorts) of patients determined, in real time, by their characteristics (i.e., patients who report that they have a new diagnosis of diabetes automatically join the specific cohort and receive online questionnaires related to diabetes).

For specific studies nested in the COOP' e-cohort, in addition to patient-reported data, for specific subgroups of patients, data collection may be expanded with the collection of biological samples to characterize the expression of chronic diseases from genomic, epigenetic, proteomic, transcriptomic and metabolomic perspectives.

Passive data enrichment

All patient-reported data are enriched by linkage at the individual level with external sources such as medico-administrative databases, hospital data warehouses, existing research registries or patients' own wearable devices. In ComPaRe, our enrichment strategy entails the systematic linkage of patient data with the following:

- Data from the French national health insurance system (SNDS), which covers about 98.8% of the French population from birth (or immigration) to death (or emigration). This database aggregates data from 1) the "Système National d'Information Inter Régimes de l'Assurance Maladie" (SNIIRAM), which includes exhaustive nationwide data on all healthcare encounters (physician or paramedical visits), medicines, medical

devices, and lab tests (without results) reimbursed to patients; 2) the “”Programme de médicalisation des systèmes d'information” (PMSI), which includes exhaustive nation-wide data on all hospitalizations in France (International Statistical Classification of Diseases and Related Health Problems, 10th revision, codes for primary and associated diagnoses, date and duration, procedures, diagnostic-related groups, and cost coding); 3) the Centre d'épidémiologie sur les causes médicales de Décès (Cépi-DC), which includes exhaustive nationwide data on the date and cause of death for the French population.

- The Assistance Publique-Hôpitaux de Paris hospital data warehouses, which contain routinely collected care data (electronic health records, medical imaging data, pathology data and laboratory test results) for all patients hospitalized in the 39 university hospitals of the Paris region (about 8.8 M patients).

Advantages of such enrichment strategy are as follows: 1) the passive, accurate and no-burden collection of objective care data, thus restricting active data collection solely to essential PROMs and PREMs; 2) the minimization of attrition bias in cohorts (even if patients no longer respond to online questionnaires, researchers are still able to know about their care consumption, hospitalizations and vital status); and 3) the minimization of the number of missing important covariates when exploiting administrative databases for research purposes (here, researchers can fill in the gaps in important confounding variables absent from these datasets by using active data collection [37]).

Creating a one-stop shop for research participation

The COOP’ e-cohort represents a simple and identifiable entry point to participate in (multiple) research projects. For researchers, it eases the identification and recruitment of patients motivated to participate in their research. For patients, it facilitates 1) the identification and participation in research projects relevant to their interests and 2) the management of their data

and consent across all research projects they participate in. For example, in ComPaRe, we propose a dynamic consent system whereby patients can easily see and modify their consent to 1) participate in the cohort, receive and answer online questionnaires, which is mandatory for participation; 2) link their data to external data (e.g., administrative or hospital databases); 3) be contacted for additional projects nested in the e-cohort and for their permission to be randomly selected to be approached for experimental interventions or to serve as control without further notice during participation in the e-cohort [38]; and 4) participate in specific nested studies.

A platform approach for research is not fundamentally new. Traditional cohorts have often been used as platform projects, with broad scientific objectives, intended to host ancillary projects. Yet, data re-use and sharing (with researchers outside of the original team) was not often anticipated or clear for participants. The COOP' e-cohort changes this by giving participants clear information about data sharing and reuse, with a large number of researchers (not yet known), in order to answer a large number of research questions (not yet identified). In fact, the COOP' e-cohort aims at transposing in epidemiology and clinical research the model of sharing common large infrastructures for accelerated research. This model is the norm rather than the exception in most scientific fields (e.g., telescopes in astronomy, colliders in physics, –omics platform in biology).

Extending to other types of research

Last but not least, the COOP' e-cohort can also accelerate research "beyond its walls" by being a reservoir of patients willing to participate in research. This model resembles the Army of Women, a community of women willing to participate in research on breast cancer. Proposals submitted to the Army of Women and approved by the scientific committee are advertised in the community and women self-select and sign up for these studies. Since 2008, the Army of Women has accelerated 120 studies with 102,000 volunteers [39, 40]. With a similar model,

COOP' e-cohorts could jumpstart projects in not only clinical research but also basic and social sciences. For example, geneticists would be able to identify type 2 diabetes patients for whom both parents also had type 2 diabetes via a cross-sectional Web survey and invite them and a control population to participate in a genetic marker study outside of the e-cohort. Then, they could decide to collaborate with epidemiologists to link the collected data with long-term outcomes and response to treatment of patients, collected from patient-reported outcomes and administrative databases.

Potential challenges and solutions for COOP' e-cohorts

First, similar to all prospective cohorts, COOP' e-cohorts are threatened by missing follow-up data and attrition bias. However, this is mitigated by 1) linkage with administrative data (even if patients no longer respond to online questionnaires; important outcomes such as their care consumption, hospitalizations and vital status are still obtained by the passive enrichment of data with administrative data sources); and 2) the involvement of patients in the selection, design and analysis of studies conducted within the cohort, which has been shown to increase engagement in research [23, 41].

Regarding quality of data, online questionnaires allow for automatic checks in the data (including cross checks between data entered in multiple questionnaires). Forms are complete. There are no data entry errors, Furthermore, linkage with administrative data allows for verifying the reliability of the information self-reported by patients. In the literature, use of online questionnaires has shown similar data quality as the use of paper questionnaires [42], interviews by skilled interviewers [43], or with data obtained from administrative databases[27, 44].

Regarding non-representativeness, online studies are known to recruit younger, more educated and more often female participants [27]. As a result, COOP' e-cohorts are not a suitable

resource for deriving generalizable disease prevalence and incidence rates. However, they are appropriate for other purposes (e.g., analytical epidemiology, development of novel patient-reported outcomes, interventional studies, etc.). In fact, the experience of the Nutrinet-Santé e-cohort and the first data collected in ComPaRe have shown that the direct outreach strategy allows for recruiting participants with diverse sociodemographic backgrounds, including socioeconomically disadvantaged individuals who are usually difficult to reach and retain in long-term epidemiologic studies [45].

Finally, concerning regulatory approval, our experience with the ComPaRe e-cohort in France has been well received from all regulatory actors in France. In fact, using data from a cohort to answer multiple research questions was a similar problem as the use of hospital data for research purposes, and COOP' e-cohorts were therefore in line with the definition of "Entrepôt de données" (data warehouse).

Discussion

In this article, we propose a new infrastructure for clinical research, the COOP' e-cohort, with the following features:

- A model for patient-centered research
- A model facilitating the reusability of tools and data: data are collected once for multiple purposes
- A model to answer questions currently not tackled by "classical" clinical research
- Flexibility and rapid data collection from e-epidemiology
- Common strata for all clinical research projects simplifying the comparison or synthesis of data
- A model reducing the incremental cost of studies and the time to set up new projects

The COOP' e-cohort offers some advantages in comparison to other existing infrastructures aiming at accelerating clinical research. In contrast to registries and registry trials, it offers extreme flexibility in the data collected and does not require the pre-existence of a registry for a given disease [46]. A platform cohort may also be easier to set up than a large international prospective clinical research network that uses hospital information systems as the main source of data [47]. Indeed, it is simpler to "import" data from different international health information systems within an existing e-cohort (in which specific data are purposefully collected) than harmonizing all information systems from hospitals, internationally (and to develop ad-hoc specific data-collection forms). Finally, different from patient communities such as PatientsLikeMe, data are collected regularly for all participants, which allows for robust longitudinal studies [48].

This new infrastructure is a tool for not just researchers; it is also designed to be a motivator to "re"-put patients at the centre of research. Thus, besides the structural advantages we present, the COOP' e-cohort may also have the power to improve transparency and quality in research because all patients involved will have the most interest in ensuring that research is rigorous and complete.

Some challenges will need to be addressed. Especially, the current academic funding systems is challenged. Indeed, the current systems rely on a project-by-project funding that is incompatible with a shared and common infrastructure for research. It will be necessary to imagine a participative research model in which researchers using the infrastructure devote a fraction of their funds for its maintenance. Also, current models of research governance will be challenged. How to make a patient democracy work? Should we look for representative patients or use regular large polls to gain the opinion of patients?

To conclude, we propose to move from a model of clinical research involving myriad small independent studies, each requiring specific funding, recruitment and data collection, to a large collaborative infrastructure of research able to combine data from different sources to answer any research question.

Acknowledgments

The authors thank Aurélie Malon, Isabelle Pane, Elise Diard, Solene Gouesbet, Carolina Riveros, Nicole Pitcher, Razak Ellafi and all members of the ComPaRe team; Laura Smales (BioMedEditing, Toronto, Canada) for editing, and Elise Diard for help in drafting figures.

Funding

ComPaRe was funded by the Assistance Publique-Hôpitaux de Paris (APHP) and Paris Descartes University

Competing interests

The authors declare no competing interests and no financial associations that may be relevant or seen as relevant to the submitted manuscript.

The authors have no association with commercial entities that could be viewed as having an interest in the general area of the submitted manuscript.

References

1. Institute of Medicine Forum on Drug Discovery D, Translation. Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary. Washington (DC): National Academies Press (US); National Academy of Sciences.; 2010.
2. Selby JV, Krumholz HM, Kuntz RE, Collins FS. Network news: powering clinical research. Science translational medicine. 2013;5(182):182fs13. Epub 2013/04/25. doi: 10.1126/scitranslmed.3006298. PubMed PMID: 23612587; PubMed Central PMCID: PMCPmc5101932.
3. Getz K, Campo R, Kaitin K. Variability in Protocol Design Complexity by Phase and Therapeutic Area Drug Information Journal. 2011; 45(4):413–20. doi: doi:10.1177/009286151104500403.
4. Getz K. Clinical Trial Complexity: The National Academies of Sciences, Engineering and Medicine; 2012. Available from: http://www.nationalacademies.org/hmd/~/media/34D1A23404A8492998AD2DF0CB6CD4D1.ashx.
5. van Staa TP, Klungel O, Smeeth L. Use of electronic healthcare records in large-scale simple randomized trials at the point of care for the documentation of value-based medicine. Journal of internal medicine. 2014;275(6):562-9. Epub 2014/03/19. doi: 10.1111/joim.12211. PubMed PMID: 24635449.
6. Coorevits P, Sundgren M, Klein GO, Bahr A, Claerhout B, Daniel C, et al. Electronic health records: new opportunities for clinical research. Journal of internal medicine. 2013;274(6):547-60. Epub 2013/08/21. doi: 10.1111/joim.12119. PubMed PMID: 23952476.
7. O'Leary E, Seow H, Julian J, Levine M, Pond GR. Data collection in cancer clinical trials: Too much of a good thing? Clinical trials (London, England). 2013;10(4):624-32. Epub 2013/06/21. doi: 10.1177/1740774513491337. PubMed PMID: 23785066.

8. Harel O, Nazzaro V. Brief Report: Rethinking Data Collection for HIV Prevention Trials. Journal of acquired immune deficiency syndromes (1999). 2016;72(4):462-4. Epub 2016/04/01. doi: 10.1097/qai.0000000000001005. PubMed PMID: 27031292; PubMed Central PMCID: PMCPmc4927344.
9. Naidoo N, Nguyen VT, Ravaud P, Young B, Amiel P, Schante D, et al. The research burden of randomized controlled trial participation: a systematic thematic synthesis of qualitative evidence. BMC medicine. 2020;18(1):6. Epub 2020/01/21. doi: 10.1186/s12916-019-1476-5. PubMed PMID: 31955710; PubMed Central PMCID: PMCPMC6970283.
10. Lazzaro C, Aloe-Spiriti M, Voso M, Latagliata R, Buccisano F, Fenu S, et al. Cost Description on a Cohort of 659 Patients with Adult MDS Included into the Italian Lazio Region Registry (the GROM-L). Blood. 2015;126(23):5237.
11. Manolio TA, Collins R. Enhancing the feasibility of large cohort studies. Jama. 2010;304(20):2290-1. Epub 2010/11/26. doi: 10.1001/jama.2010.1686. PubMed PMID: 21098774; PubMed Central PMCID: PMCPmc3075846.
12. Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, et al. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. Health technology assessment (Winchester, England). 2007;11(48):iii, ix-105. Epub 2007/11/15. PubMed PMID: 17999843.
13. Dechartres A, Riveros C, Harroch M, Faber T, Ravaud P. Characteristics and Public Availability of Results of Clinical Trials on Rare Diseases Registered at Clinicaltrials.gov. JAMA Intern Med. 2016;176(4):556-8. Epub 2016/03/10. doi: 10.1001/jamainternmed.2016.0137. PubMed PMID: 26954340.
14. Ioannidis JP. Are mortality differences detected by administrative data reliable and actionable? Jama. 2013;309(13):1410-1. Epub 2013/04/04. doi: 10.1001/jama.2013.3150. PubMed PMID: 23549588.
15. Kaplan CP, Napoles AM, Dohan D, Shelley Hwang E, Melisko M, Nickleach D, et al. Clinical trial discussion, referral, and recruitment: physician, patient, and system factors. Cancer causes & control : CCC. 2013;24(5):979-88. Epub 2013/02/20. doi: 10.1007/s10552-013-0173-5. PubMed PMID: 23420328.
16. Pestine E, Stokes A, Trinquart L. Representation of obese participants in obesity-related cancer randomized trials. Annals of oncology : official journal of the European Society for Medical Oncology. 2018;29(7):1582-7. Epub 2018/06/14. doi: 10.1093/annonc/mdy138. PubMed PMID: 29897392.
17. Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. Jama. 2007;297(11):1233-40. Epub 2007/03/22. doi: 10.1001/jama.297.11.1233. PubMed PMID: 17374817.
18. Buffel du Vaure C, Dechartres A, Battin C, Ravaud P, Boutron I. Exclusion of patients with concomitant chronic conditions in ongoing randomised controlled trials targeting 10 common chronic conditions and registered at ClinicalTrials.gov: a systematic review of registration details. BMJ open. 2016;6(9):e012265. Epub 2016/09/30. doi: 10.1136/bmjopen-2016-012265. PubMed PMID: 27678540; PubMed Central PMCID: PMCPMC5051474.
19. Tallon D, Chard J, Dieppe P. Relation between agendas of the research community and the research consumer. Lancet (London, England). 2000;355(9220):2037-40. Epub 2000/07/08. doi: 10.1016/s0140-6736(00)02351-5. PubMed PMID: 10885355.
20. Gandhi GY, Murad MH, Fujiyoshi A, Mullan RJ, Flynn DN, Elamin MB, et al. Patient-important outcomes in registered diabetes trials. Jama. 2008;299(21):2543-9. Epub 2008/06/05. doi: 10.1001/jama.299.21.2543. PubMed PMID: 18523223.
21. Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, et al. Increasing value and reducing waste in biomedical research regulation and management. Lancet (London, England). 2014;383(9912):176-85. Epub 2014/01/15. doi: 10.1016/s0140-6736(13)62297-7. PubMed PMID: 24411646; PubMed Central PMCID: PMCPMC3952153.

22. Olfson M, Wall MM, Blanco C. Incentivizing Data Sharing and Collaboration in Medical Research-The S-Index. JAMA psychiatry. 2017;74(1):5-6. Epub 2016/10/27. doi: 10.1001/jamapsychiatry.2016.2610. PubMed PMID: 27784040.
23. Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. Nature biotechnology. 2011;29(5):411-4. Epub 2011/04/26. doi: 10.1038/nbt.1837. PubMed PMID: 21516084.
24. Kaye J, Curren L, Anderson N, Edwards K, Fullerton SM, Kanellopoulou N, et al. From patients to partners: participant-centric initiatives in biomedical research. Nature reviews Genetics. 2012;13(5):371-6. Epub 2012/04/05. doi: 10.1038/nrg3218. PubMed PMID: 22473380; PubMed Central PMCID: PMCPmc3806497.
25. The best research is produced when researchers and communities work together. Nature. 2018;562(7725):7. Epub 2018/10/05. doi: 10.1038/d41586-018-06855-7. PubMed PMID: 30283119.
26. National Institutes of Health. All of Us Research Protocol. National Institutes of Health, 2018.
27. Andreeva VA, Salanave B, Castetbon K, Deschamps V, Vernay M, Kesse-Guyot E, et al. Comparison of the sociodemographic characteristics of the large NutriNet-Sante e-cohort with French Census data: the issue of volunteer bias revisited. Journal of epidemiology and community health. 2015;69(9):893-8. Epub 2015/04/03. doi: 10.1136/jech-2014-205263. PubMed PMID: 25832451.
28. Collins R. What makes UK Biobank special? Lancet (London, England). 2012;379(9822):1173-4. Epub 2012/04/03. doi: 10.1016/s0140-6736(12)60404-8. PubMed PMID: 22463865.
29. Hercberg S, Castetbon K, Czernichow S, Malon A, Mejean C, Kesse E, et al. The Nutrinet-Sante Study: a web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status. BMC public health. 2010;10:242. Epub 2010/05/13. doi: 10.1186/1471-2458-10-242. PubMed PMID: 20459807; PubMed Central PMCID: PMC2881098.
30. Tran VT, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. NPJ Digit Med. 2019;2:53. Epub 2019/07/16. doi: 10.1038/s41746-019-0132-y. PubMed PMID: 31304399; PubMed Central PMCID: PMCPMC6572821.
31. Tran VT, Barnes C, Montori VM, Falissard B, Ravaud P. Taxonomy of the burden of treatment: a multicountry Web-based qualitative study of patients with chronic conditions. BMC medicine. 2015;13:115. doi: 10.1186/s12916-015-0356-x.
32. Tran VT, Riveros C, Pean C, Czarnobroda A, Ravaud P. Patients' perspective on how to improve the care of people with chronic conditions in France: a citizen science study within the ComPaRe e-cohort. BMJ quality & safety. 2019. Epub 2019/04/25. doi: 10.1136/bmjqs-2018-008593. PubMed PMID: 31015376.
33. Tran VT, Montori VM, Ravaud P. Is My Patient Overwhelmed?: Determining Thresholds for Acceptable Burden of Treatment Using Data From the ComPaRe e-Cohort. Mayo Clinic proceedings. 2019. Epub 2019/10/18. doi: 10.1016/j.mayocp.2019.09.004. PubMed PMID: 31619365.
34. Relton C, Torgerson D, O'Cathain A, Nicholl J. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. BMJ (Clinical research ed). 2010;340:c1066. Epub 2010/03/23. PubMed PMID: 20304934.
35. Perrier L, Blondal E, MacDonald H. The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. PloS one. 2020;15(2):e0229182. Epub 2020/02/28. doi: 10.1371/journal.pone.0229182. PubMed PMID: 32106224; PubMed Central PMCID: PMCPMC7046208.
36. Sertkaya A, Wong HH, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. Clinical trials (London, England). 2016;13(2):117-26. Epub 2016/02/26. doi: 10.1177/1740774515625964. PubMed PMID: 26908540.
37. Ioannidis JP. Are mortality differences detected by administrative data reliable and actionable? (1538-3598 (Electronic)).
38. Young-Afat DA, Verkooijen HA, van Gils CH, van der Velden JM, Burbach JP, Elias SG, et al. Brief Report: Staged-informed Consent in the Cohort Multiple Randomized Controlled Trial Design.

Epidemiology (Cambridge, Mass). 2016;27(3):389-92. Epub 2016/04/02. doi: 10.1097/ede.0000000000000435. PubMed PMID: 27035689.
39. Dr Susan Love Research Foundation. Army of Women: Dr Susan Love Research Foundation; 2018 [cited 2019 02/01]. Available from: https://www.armyofwomen.org/.
40. Love S, Mills D, Eraklis E, Hurlbert M. The Love/Avon Army of Women – Accelerating Research into the Etiology and Prevention of Breast Cancer. Thirty-Second Annual CTRC-AACR San Antonio Breast Cancer Symposium; San Antonio, TX2009.
41. Wicks P, Heywood B, Heywood J. Online platform to accelerate patient involvement in open instrument development. Quality of Life Research. 2013;22(1):55.
42. Touvier M, Mejean C, Kesse-Guyot E, Pollet C, Malon A, Castetbon K, et al. Comparison between web-based and paper versions of a self-administered anthropometric questionnaire. European journal of epidemiology. 2010;25(5):287-96. Epub 2010/03/02. doi: 10.1007/s10654-010-9433-9. PubMed PMID: 20191377.
43. Touvier M, Kesse-Guyot E, Mejean C, Pollet C, Malon A, Castetbon K, et al. Comparison between an interactive web-based self-administered 24 h dietary record and an interview by a dietitian for large-scale epidemiological studies. The British journal of nutrition. 2011;105(7):1055-64. Epub 2010/11/18. doi: 10.1017/s0007114510004617. PubMed PMID: 21080983.
44. Andreeva VA, Deschamps V, Salanave B, Castetbon K, Verdot C, Kesse-Guyot E, et al. Comparison of Dietary Intakes Between a Large Online Cohort Study (Etude NutriNet-Sante) and a Nationally Representative Cross-Sectional Study (Etude Nationale Nutrition Sante) in France: Addressing the Issue of Generalizability in E-Epidemiology. American journal of epidemiology. 2016;184(9):660-9. Epub 2016/11/03. doi: 10.1093/aje/kww016. PubMed PMID: 27744386; PubMed Central PMCID: PMCPmc5101865.
45. Kesse-Guyot E, Andreeva V, Castetbon K, Vernay M, Touvier M, Mejean C, et al. Participant profiles according to recruitment source in a large Web-based prospective study: experience from the Nutrinet-Sante study. Journal of medical Internet research. 2013;15(9):e205. Epub 2013/09/17. doi: 10.2196/jmir.2488. PubMed PMID: 24036068; PubMed Central PMCID: PMCPmc3785981.
46. Lauer MS, D'Agostino RB, Sr. The randomized registry trial--the next disruptive technology in clinical research? The New England journal of medicine. 2013;369(17):1579-81. Epub 2013/09/03. doi: 10.1056/NEJMp1310102. PubMed PMID: 23991657.
47. Rachel R, Richard P, Gregory S, Lesley C, Reesa L, Adrian H, et al. Using Electronic Health Record Data in Pragmatic Clinical Trials: NIH HCS collaboratory; 2015 [cited 2018 9/11]. Available from: http://rethinkingclinicaltrials.org/chapters/design/using-electronic-health-record-data-pragmatic-clinical-trials-top/using-electronic-health-record-data-in-pragmatic-clinical-trials-introduction/.
48. van der Eijk M, Faber MJ, Aarts JW, Kremer JA, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. Journal of medical Internet research. 2013;15(6):e115. Epub 2013/06/28. doi: 10.2196/jmir.2476. PubMed PMID: 23803284; PubMed Central PMCID: PMCPmc3713879.
49. Johnson JA, Coons SJ, Ergo A, Szava-Kovats G. Valuation of EuroQOL (EQ-5D) health states in an adult US sample. PharmacoEconomics. 1998;13(4):421-33. Epub 1998/03/08. PubMed PMID: 10178666.
50. Tran VT, Montori VM, Eton DT, Baruch D, Falissard B, Ravaud P. Development and description of measurement properties of an instrument to assess treatment burden among patients with multiple chronic conditions. BMC medicine. 2012;10:68. Epub 2012/07/06. doi: 10.1186/1741-7015-10-68. PubMed PMID: 22762722; PubMed Central PMCID: PMCPmc3402984.
51. Sidorkiewicz S, Tran VT, Cousyn C, Perrodeau E, Ravaud P. Development and validation of an instrument to assess treatment adherence for each individual drug taken by a patient. BMJ open. 2016;6(5):e010510. Epub 2016/05/12. doi: 10.1136/bmjopen-2015-010510. PubMed PMID: 27165645; PubMed Central PMCID: PMCPMC4874131.
52. Paterson C. Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. BMJ (Clinical research ed). 1996;312(7037):1016-20. Epub 1996/04/20. PubMed PMID: 8616351; PubMed Central PMCID: PMC2350813.

53. Salzes C, Abadie S, Seneschal J, Whitton M, Meurant JM, Jouary T, et al. The Vitiligo Impact Patient Scale (VIPs): Development and Validation of a Vitiligo Burden Assessment Tool. The Journal of investigative dermatology. 2016;136(1):52-8. Epub 2016/01/15. doi: 10.1038/jid.2015.398. PubMed PMID: 26763423.
54. Welch GW, Jacobson AM, Polonsky WH. The Problem Areas in Diabetes Scale. An evaluation of its clinical utility. Diabetes care. 1997;20(5):760-6. Epub 1997/05/01. PubMed PMID: 9135939.

List of boxes and tables

Box 1: Objectives of the COOP' e-cohort

Box 2: Community of Patients for Research (ComPaRe)

Table: Example of studies and design that could be conducted with COOP' e-cohorts

List of figures

Figure 1: COOP' e-cohorts to accelerate research. Duration and delays with a classical pen-and-paper study are from a study conducted by our team in France (Tran VT, BMC Med 2012). Duration and delays for a study nested in the COOP' e-cohort are from a study conducted in ComPaRe in France (Tran VT, npj Digit Med 2019).

Figure 2: Recruitment and data collection in a COOP' e-cohort

Box 1: Objectives of the COOP' e-cohort

The COOP' e-cohort uses a shared infrastructure for research for the following aims:

- Accelerate research within and outside of the cohort
- Reduce the costs of research
- Facilitate regulatory, procedural and logistics efforts required for research
- Reduce waste of research via the mutualization of data
- Reduce the burden of research for participants
- Forego silos in research and encourage multidisciplinary projects

Box 2: Community of Patients for Research (ComPaRe)

ComPaRe (https://compare.aphp.fr) is a proof of concept for a COllaborative Open Platform (COOP') e-cohort in the field of chronic conditions.
Patients join ComPaRe to donate some of their time to accelerate research on chronic conditions. They answer regular online questionnaires (using patient-reported outcome measures [PROMs], patient-reported experience measures [PREMs] or questionnaires with open-ended questions), suggest ideas for new research, and can participate in the set-up or analysis of research projects.
For researchers, ComPaRe offers the ability to quickly set up new projects with PROMs or PREMs in the field of chronic conditions.

Aim:

- To evaluate the impact of chronic conditions and their management on patients' lives
- To answer various research questions on specific chronic conditions or treatments
- To evaluate the importance of multi-morbidity and its consequences for patients.

Expected number of participants: 100 000 (10,000 per year) over 10 years.
Currently, 36,000 participants have joined the project since January 2017.

Eligibility to enter ComPaRe: Eligible participants are adults ≥ 18 years old with at least one chronic condition defined as a condition requiring regular health care for at least 6 months.

Recruitment strategy: Patients are informed that the project exists by the following:

- Partner researchers, physicians and national medical societies
- Systematic invitation of people who 1) booked their medical appointment online in partner hospitals and 2) consented to receive information on research (approximately 500 participants per month)
- Advertisement on social networks (approximately 2000 followers on various social media)
- A general media campaign (TV, radio, newspapers)
- Snowball sampling method

Characteristics measured at inclusion:

- Socio-demographic characteristics
- Chronic conditions (including year of diagnosis) based on chronic conditions listed in the International Statistical Classification of Diseases and Related Health Problems 10th version, and International Classification of Primary Care, 2nd edition
- Long-term treatments for patients (start/end date, dosage)
- Disabilities
- Anthropometric data: self-measured weight and height
- Tobacco use, alcohol consumption

Follow-up with patient-reported outcomes:

- Example of PROMs collected regularly for all patients in the cohort:
  - Quality of life using the EQ5D [49]
  - Treatment burden using the Treatment Burden Questionnaire [50]
  - Adherence to medications for each drug taken by the patient [51]
  - Patients' self-evaluation of their symptoms with the Measure Your Medical Outcome Profile [52]
- Example of PROMs used in specific "sub-cohorts":
  - Vitiligo cohort: severity of disease with the Vitiligo Impact Patient Scale [53]
  - Diabetes cohort: problem areas in diabetes with the Problem Areas in Diabetes questionnaire [54]

Enrichment of patient-reported data with other data sources (planned)

- Linkage with the French National Health System databases (SNDS), which covers data on hospital stays, patients' diagnoses and reimbursement of medications (characterized by their Anatomical Therapeutic Chemical code), medical procedures, laboratory tests and medical or paramedical visits and information on participants' vital status
- Linkage with hospital data warehouses (electronic health records, biological and imaging data) for patients receiving care in the Paris region
- Linkage with patients' own wearable devices (smartphones, smart-watches, etc.)
- Specific collection of biological data

Patients' participation in research:

- 44% of participants want to do more than simply answer online questionnaires. They are willing to test questionnaires, participate in data analysis, be active in the recruitment of new participants etc.
- 1500 participants have contributed to INSPIRE and have proposed at least one research idea, for a total of 2000 research ideas.

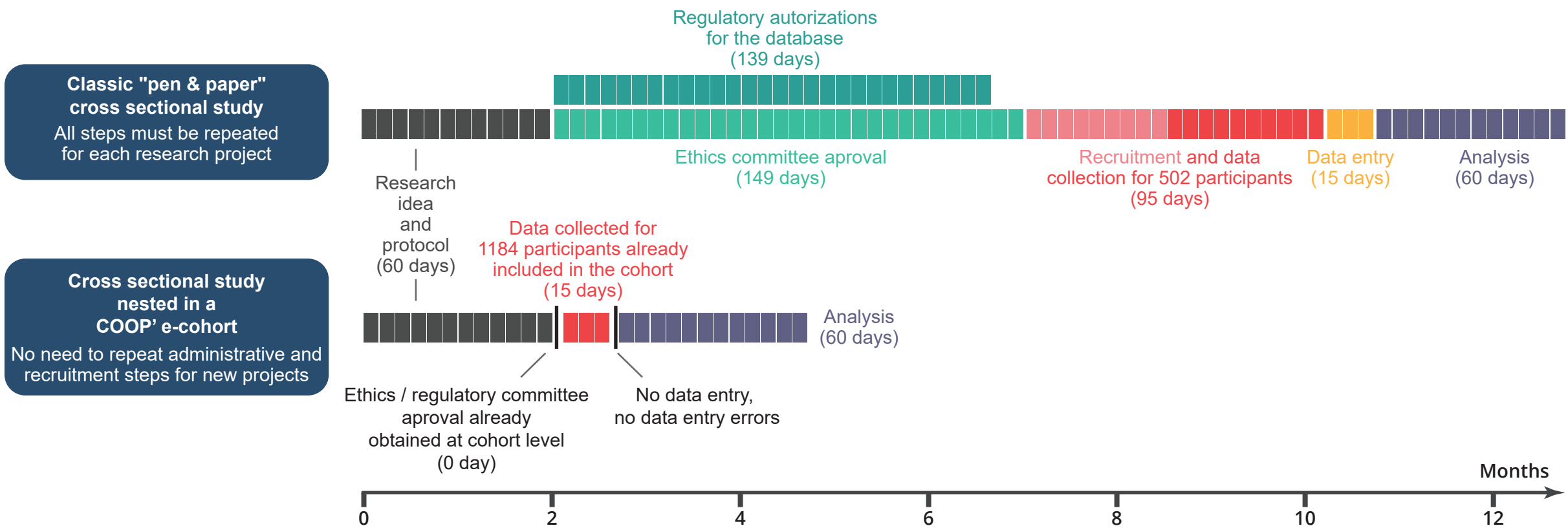
Example of projects that are possible by using ComPaRe

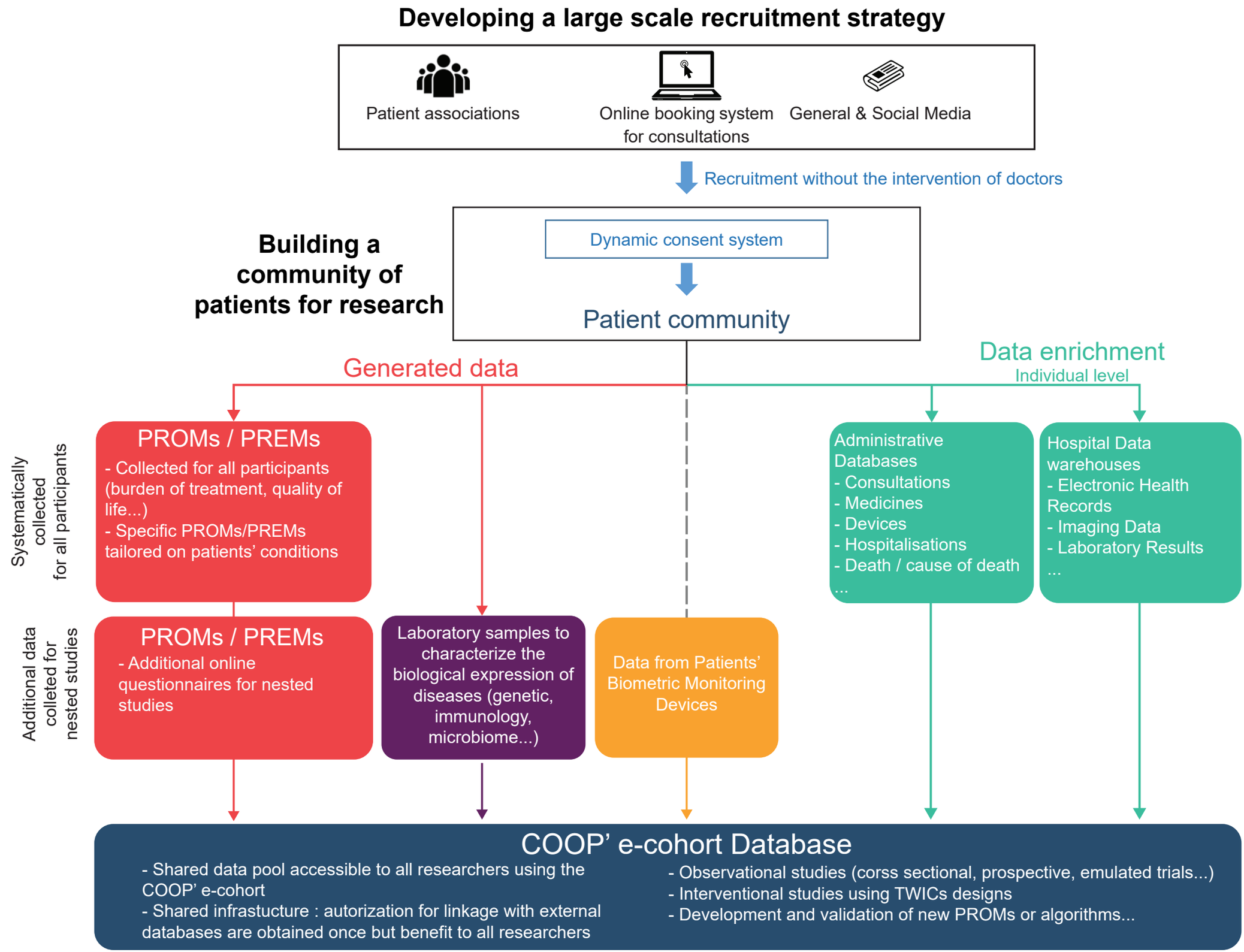
- *Help for recruitment* (in clinical or fundamental research or social sciences etc.). Potential eligible participants can be targeted by using all data collected in the cohort and invited by e-mail for studies outside of ComPaRe.
- *Observational studies* (cross-sectional or longitudinal). We evaluated the proportion of patients who considered their burden of treatment unacceptable [33].
- *Online citizen science studies*. We engaged patients with chronic conditions to generate ideas on how to improve their care by using an open-ended question "If you had a magic wand, what would you change in your care?" [32]
- *Development of new PROMs*. We are developing a new measure to assess the burden of treatment for Vitiligo.
- *Nested randomized trials*. PaNaM is a cohort multiple randomised controlled trial aimed at evaluating the impact of a virtual patient navigator on mortality, hospitalization rates and burden of treatment for multi-morbid patients

Table: Example of studies and design that can be conducted within COOP' e-cohorts

Studies/designs that can be conducted within COOP' e-cohorts	**Fields/domains for which COOP' e-cohorts could support recruitment**

- Analytical epidemiology aimed at inferring associations between exposures and outcomes using observational studies (cross-sectional, prospective, retrospective) relying mainly on PROM and PREM data
 	
- Comparative effectiveness research studies such as:
  o Emulated trials in which interventions are defined by data from medico-administrative or hospital databases and evaluation involves using PROMs
  o Trials assessing online interventions (e.g., educational videos, online educative programs) assessed by using PROMs
  o Bring Your Own Device (BYOD) trials in which the intervention is, for example, a smartphone app to be downloaded and is assessed by using PROMs
 	
- Development of new PROMs and PREMs
- Development of predictive models or algorithms based on PROMs/PREMs
 	
- Online surveys with open-ended questions aimed at exploring new concepts
- Online collective intelligence studies aimed at generating ideas on a given concept
	- Comparative effectiveness research studies aimed at evaluating drugs/devices in trials
	
	- Clinical research requiring face-to-face contact and/or biological samples
	
	- Identifying participants for basic science studies (genetics, immunology, etc.)
	
	- Social science studies involving patients with chronic conditions


Classic "pen & paper" cross sectional study
All steps must be repeated for each research project
Regulatory autorizations for the database (139 days)
Ethics committee aproval (149 days)
Recruitment and data collection for 502 participants (95 days)
Data entry (15 days)
Analysis (60 days)
Research idea and protocol (60 days)
Cross sectional study nested in a COOP' e-cohort
No need to repeat administrative and recruitment steps for new projects
Data collected for 1184 participants already included in the cohort (15 days)
Analysis (60 days)
Ethics / regulatory committee aproval already obtained at cohort level (0 day)
No data entry, no data entry errors
Months
0
2
4
6
8
10
12


Developing a large scale recruitment strategy
Patient associations
Online booking system for consultations
General & Social Media
Recruitment without the intervention of doctors
Building a community of patients for research
Dynamic consent system
Patient community
Generated data
Data enrichment
Individual level
Systematically collected for all participants
PROMs / PREMs
- Collected for all participants (burden of treatment, quality of life...)
- Specific PROMs/PREMs tailored on patients' conditions
Administrative Databases
- Consultations
- Medicines
- Devices
- Hospitalisations
- Death / cause of death
...
Hospital Data warehouses
- Electronic Health Records
- Imaging Data
- Laboratory Results
...
Additional data colleted for nested studies
PROMs / PREMs
- Additional online questionnaires for nested studies
Laboratory samples to characterize the biological expression of diseases (genetic, immunology, microbiome...)
Data from Patients' Biometric Monitoring Devices
COOP' e-cohort Database
- Shared data pool accessible to all researchers using the COOP' e-cohort
- Shared infrastucture : autorization for linkage with external databases are obtained once but benefit to all researchers
- Observational studies (corss sectional, prospective, emulated trials...)
- Interventional studies using TWICs designs
- Development and validation of new PROMs or algorithms...