eTRANSAFE: data science to empower translational safety assessment

Ferran Sanz, François Pognan, Thomas Steger-Hartmann, Carlos Díaz and the eTRANSAFE Consortium*

Thirteen pharmaceutical companies have shared and integrated preclinical and clinical data for creating computational resources that enhance translational drug safety assessment.

Drug research and development (R&D) is knowledge-intensive, and so can benefit from advances in data science and technologies, such as automated data curation, data integration and artificial intelligence. Data sharing beyond the boundaries of individual pharmaceutical companies offers particularly interesting opportunities for accelerating and improving aspects of the R&D process. However, balancing collaboration and confidentiality remains a complex challenge for the industry.

Drug safety is not the main domain for competition between pharmaceutical companies, and so it is more feasible to develop collaborative projects in this area. For example, legacy drug safety data may be shared among companies with the aim of improving drug safety assessment. Significant steps forward in this direction were taken by the eTOX and eTRANSAFE projects. The eTOX project, which focused on sharing preclinical toxicological data, was carried out from 2010 to 2017 (see Related links). More recently, the eTRANSAFE project, addressing translational drug safety assessment through integrative knowledge management[1], has been running from 2017 to 2023 (see Related links). Both projects are part of the Innovative Medicines Initiative (IMI) funded by the European Commission and the European Federation of Pharmaceutical Industries and Associations (EFPIA). Here, we highlight the achievements of the eTRANSAFE initiative and discuss the impact of the tools it has developed.

Achievements of the eTRANSAFE initiative

Extensions in data-sharing. Although both eTOX and eTRANSAFE were based on the potential of sharing legacy data for drug safety assessment, eTRANSAFE incorporates substantial advances compared with the eTOX project. Only aggregated data were shared in eTOX, whereas repeated dose toxicology (RDT) studies at the individual animal level of detail have been shared in eTRANSAFE. These RDT data are coded in the CDISC Standard for Exchange of Nonclinical Data (SEND) and they are electronically extracted from the laboratory information management systems of the companies, avoiding human transcription errors. Almost 10,000 RDT studies, not available anywhere, have been shared by the companies in eTRANSAFE. Non-SEND historical data, like those previously shared in eTOX, have been mapped to SEND to maximise standardized RDT data available in eTRANSAFE.

Incorporation of new data types. Another advance in comparison to eTOX has been the extraction of treatment-related findings from the free-text RDT study reports by means of a natural language processing tool developed in the project, as well as a SEND-adapted 'study report domain template and editor' created for incorporating this information into the eTRANSAFE preclinical database. Moreover, pharmaceutical companies have also exchanged pharmacovigilance data extracted from their periodic safety update reports (PSURs), as well as in vitro off-target safety pharmacology data, which have been used to create new legacy databases.

Currently, eTRANSAFE integrates information from 12 databases that contain proprietary and publicly available data generated in preclinical and clinical research (eTRANSAFE preclinical database, eTOX database, off-target pharmacology database, ChEMBL, a database of drugs withdrawn due to safety concerns, a database of drug adverse effects extracted from ClinicalTrials.gov, a database of biomarkers extracted by text mining from ClinicalTrials.gov, DailyMed, a database on drug adverse events from publications in Medline, FAERS, FDA DILIrank, and a database of periodic safety update reports). The inclusion of clinical databases provided opportunities for translational analysis; that is, investigating the concordance between the results obtained in preclinical studies and the safety observations in clinical studies. The project has also developed new tools for exploring knowledge about the molecular basis of toxicological events[2].

Development of a platform for data integration and analysis. The eTRANSAFE project has developed a modular software platform —ToxHub — that allows the integrated exploration and exploitation of all the above-mentioned databases. Moreover, ToxHub incorporates a software tool for a universal and user-friendly incorporation of additional databases, a computational service enabling preclinical–clinical semantic interoperability by aligning the preclinical terminologies of SEND with the clinical vocabulary of MedDRA through SNOMED CT (eTRANSAFE Rosetta Stone), a chemistry service for identifying in an unambiguous way the structures of the drugs included in the different databases by means of SMILES representations and InChI keys, and a structural similarity service for expanding the queries with compounds showing structural similarity or containing given substructures.

ToxHub also incorporates advanced software for data visualization (Sirona) that is finely tuned to support the needs of drug safety professionals, as well as a computational platform (Flame[3]) designed for user-friendly development, documentation and use of predictive models based on machine learning algorithms. Moreover, Flame incorporates functionalities for the development of ensemble models and for the export of models without including any kind of information about the compounds used to build the training set. This constitutes a critical functionality to enable cooperative model building using data that are not sharable.

A collection of predictive models has already been developed and made available in the ToxHub, including a series of predictive models of drug–transporter interactions.

The eTRANSAFE project has just ended as an IMI-funded project, but it has clear avenues for the sustainability and exploitation of its results. First, the databases and software that have been developed will be permanently stored by one of the 'honest brokers' of the project, Universitat Pompeu Fabra, to keep them available to the project partners and to other interested parties. Also, several software developments of eTRANSAFE are distributed as open source, such as Rosetta Stone and Flame. In addition, a commercial partner (Instem) has been ap-

pointed to carry out the maintenance, commercialization and further development of the ToxHub platform.

Application and impact in pharmaceutical research

The ToxHub platform is already operational in many of the pharmaceutical companies involved in eTRANSAFE, which are using the platform in their daily practice. Several use cases have demonstrated how ToxHub improves our ability to compare preclinical and clinical drug safety data, such as a comparative analysis of the skin toxicity of kinase inhibitors, and data has been used to prototype the concept of a virtual control group for preclinical in vivo studies.

Kinase inhibitors use case. Kinase inhibitors are one of the most important drug classes in cancer treatment, but their inhibition of key processes of the cell cycle results in numerous adverse events, which are sometimes dose-limiting or even life-threatening. Skin reactions are one of the dose-limiting toxicities of kinase inhibitors. Most of the adverse events are generally identified during preclinical development in animal studies. However, systematic investigations on how far these observations translate into serious adverse events and which animal species is the most predictive for specific events are scarce.

ToxHub allows access to preclinical data of sufficient volume and granularity to perform powerful comparisons in both the preclinical and clinical domains. In this use case[5], preclinical data were searched in ToxHub for skin findings reported with kinase inhibitors. Preclinical data for more than one species were found for gefitinib, imatinib, nilotinib and vandetanib. In a second step, the clinical trials and the FAERS databases integrated into the ToxHub were searched for clinical skin-related findings for these four kinase inhibitors. Nilotinib had only few entries regarding severe skin reactions in the clinical trials database and no entries in FAERS, whereas gefitinib, imatinib and vandetanib had various entries of severe skin findings such as ulcers and dermatitis. The comparison of the preclinical and the clinical results indicated that in this case, the rat was the most sensitive species to detect severe skin findings, such as atrophy and ulcers also occurring clinically with gefitinib and imatinib, whereas non-human primates were less sensitive[5]. This higher concordance between rats and humans compared to monkeys and humans regarding skin reactions to kinase inhibitors is in accordance with an analysis previously performed for a wider range of drugs[4].

Prototyping the virtual control group concept. eTRANSAFE has also successfully prototyped a promising application involving the reuse of data from control groups obtained in legacy preclinical in vivo studies under good laboratory practice (GLP) conditions to develop virtual control groups[6]. Data from control group animals from rodent and non-rodent RDT studies were collected, curated and statistically analysed for distribution of values and the potential changes of the values over time. Based on these analyses, a best practice procedure has been drafted, which describes the steps that should be undertaken for an optimal match of the virtual controls with animals used in future studies to partially or entirely replace the concurrent control animals.

A qualification of this procedure will be undertaken by some of the eTRANSAFE partners re-assessing legacy studies in which the concurrent controls will be replaced by virtual controls. The effect of this replacement will be evaluated by considering the ability to identify the same treatment-related effects as in the original studies. Implementing this concept will still require extended collection of data from control animals, computational tools for generating virtual control animals and a thorough validation in close collaboration with regulatory authorities, with which interactions have already started. Virtual animals could in time substitute, at least in part, for the animals currently used in experimental control groups, thereby reducing the number of animals used in preclinical safety assessment, in line with the 3Rs policies on replacement, reduction and refinement.

Conclusion

The eTRANSAFE project shows how integration of preclinical and clinical data, combined with a computational toolbox customized to the needs of pharmaceutical safety experts, enables efficient data retrieval and translational analysis. It also exemplifies how modern data science can increase comprehensiveness, accuracy and speed in assessing safety issues during drug development.

Ferran Sanz[1†], François Pognan[2†], Thomas Steger-Hartmann[3†] Carlos Díaz[4†] and the eTRANSAFE Consortium*

[1]Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Spain.
[2]Novartis Pharma, Basel, Switzerland.
[3]Bayer AG, Pharmaceuticals, Berlin, Germany.
[4]Synapse Research Management Partners, Madrid, Spain.
[*]See online version for details
[†]e-mails: ferran.sanz@upf.edu; francois.pognan@novartis.com; thomas.steger-hartmann@bayer.com; cdiaz@synapse-managers.com

References

1. Pognan, F. *at al.* The eTRANSAFE project on translational safety assessment through integrative knowledge management: achievements and perspectives. *Pharmaceuticals (Basel).* **14,** 237 (2021).
2. Callegaro, G. *at al.* The human hepatocyte TXG-MAPr: gene co-expression network modules to support mechanism-based risk assessment. *Arch. Toxicol.* **95,** 3745-3775 (2021).
3. Pastor, M. *et al.* Flame: an open-source framework for model development, hosting, and usage in production environments. *J. Cheminform.* **13,** 31 (2021).
4. Clark, M. *et al.* A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans. *Regul. Toxicol. Pharmacol.* **96,** 94-105 (2018).
5. Steger-Hartmann, T. *et al.* Perspectives of data science in preclinical safety assessment. *Drug Discov. Today* XXXXXX (2023).
6. Steger-Hartmann, T. *et al*. Introducing the concept of virtual control groups into preclinical toxicology testing. *ALTEX.* **37,** 343-349 (2020).

Competing interests

Some authors are employed in the pharmaceutical industry or in small-and-medium sized enterprises, as indicated in their affiliations. The other authors declare no competing interests.

Related links

eTOX: http://www.etoxproject.eu
eTRANSAFE: https://etransafe.eu/

***Full author list**

Ferran Sanz[1†], François Pognan[2†], Thomas Steger-Hartmann[3†], Carlos Díaz[4†], Shoji Asakura[5], Alexander Amberg[6], Nathalie Bécourt-Lhote[7], Niklas Blomberg[8], Nicolas Bosc[9], Katharine Briggs[10], Frank Bringezu[11], Claire Brulle-Wohlhueter[6], Søren Brunak[12], Ruud Bueters[13], Giulia Callegaro[14], Salvador Capella[15], Emilio Centeno[1], Javier Corvi[15], Mark

T. D. Cronin[16], Philip Drew[17], Guillemette Duchateau-Nguyen[18], Gerhard F. Ecker[19], Sylvia Escher[20], Eloy Felix[9], Miguel Ferreiro[21], Markus Frericks[22], Laura I. Furlong[1,23], Robert Geiger[22], Catherine George[24], Melanie Grandits[19], Dragomir Ivanov-Draganov[18], Jean Kilgour-Christie[2], Tevfik Kiziloren[9], Jan A. Kors[25], Naoki Koyama[5], Annika Kreuchwig[3], Andrew R. Leach[9], Miguel Angel Mayer[1], Peter Monecke[6], Wolfgang Muster[18], Chihiro Miyamoto Nakazawa[5], Gavin Nicholson[26], Rowan Parry[25], Manuel Pastor[1], Janet Piñero[1,23], Nils Oberhauser[2], Juan Manuel Ramírez-Anguita[1], Adrián Rodrigo[21], Aljosa Smajic[19], Markus Schaefer[6], Sebastian Schieferdecker[27], Inari Soininen[4], Emma Terricabras[1], Panuwat Trairatphisan[28], Sean C. Turner[28], Alfonso Valencia[15], Bob van de Water[14], Johan L. van der Lei[4,25], Erik M. van Mulligen[25], Esther Vock[27], David Wilkinson[10]

[1]Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Spain.
[2]Novartis Pharma, Basel, Switzerland.
[3]Bayer, Berlin, Germany.
[4]Synapse Research Management Partners, Madrid, Spain.
[5]Eisai Co., Ltd. Tsukuba, Ibaraki, Japan.
[6]Preclinical Safety, Sanofi, 65926 Frankfurt am Main, Germany.
[7]Non-Clinical Drug Safety, Translational Medicine, Servier, 92284 Suresnes, France.
[8]ELIXIR, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK.
[9]European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK.
[10]Lhasa Limited, Leeds LS11 5PS, UK.
[11]Chemical and Preclinical Safety, Merck Healthcare KGaA, 64293 Darmstadt, Germany.
[12]Technical University of Denmark (DTU), 2800 Lyngby, Denmark (Current address: Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, 2200 Copenhagen, Denmark).
[13]Janssen Research & Development, Janssen Pharmaceutica N.V., B-2340 Beerse, Belgium.
[14]Leiden Academic Centre for Drug Research (LACDR), Leiden University, 2300 RA Leiden, The Netherlands.
[15]Barcelona Supercomputing Center, 08034 Barcelona, Spain.
[16]Liverpool John Moores University, Liverpool L3 3AF, UK.
[17]PDS Consultants, Leicester LE1 5XY, UK.
[18]Roche Pharmaceutical Research & Early Development, Pharmaceutical Sciences, Roche Innovation Center, 4070 Basel, Switzerland.
[19]Universität Wien, 1090 Vienna, Austria.
[20]Fraunhofer Institute for Toxicology and Experimental Medicine (ITEM), 30625 Hannover, Germany.
[21]GMV, 28760 Tres Cantos, Madrid, Spain.
[22]BASF, 67056 Ludwigshafen, Germany.
[23]MedBioinformatics Solutions SL, 08018 Barcelona, Spain.
[24]Ipsen Innovation, 91940 Les Ulis, France.
[25]Department of Medical Informatics, Erasmus University Medical Center, 3015 GD Rotterdam, the Netherlands.
[26]Optibrium Ltd, Cambridge Innovation Park, Cambridge CB25 9GL, UK.
[27]Boehringer Ingelheim Pharma GmbH & Co. KG, 88379 Biberach an der Riss, Germany.

[28]AbbVie Germany GmbH & Co. KG, 67061 Ludwigshafen am Rhein, Germany.

[†]e-mails: ferran.sanz@upf.edu; francois.pognan@novartis.com; thomas.steger-hartmann@bayer.com; cdiaz@synapse-managers.com