A Guide to Electronic Data Capture in Veterinary Clinical Studies

In the last decade, electronic data capture (EDC) has become an attractive and convenient alternative to conventional paper-based methods of data collection in clinical studies. A variety of different EDC solutions is available on the market, each providing its own features. To take full advantage of the features and options EDC systems offer, it requires experienced data managers and thorough planning of studies. This article will share years of experience on what to consider and potential challenges when building EDC setups for veterinary clinical studies from a data manager's perspective.

Traditionally, recording data on paper and manually transcribing them into a database has been common practice in clinical studies. In the light of an ever-increasing digitalisation in all areas of our everyday life, it is not a surprise that we have observed a strong trend towards electronic ways of data capture in clinical studies over the past years. Electronic data capture (EDC) systems are software solutions, nowadays typically web-based user interfaces with a powerful database in the background, specifically designed to collect study data in electronic form. Perhaps the most beneficial feature of EDC systems is that they provide real-time access to the data. In contrast to that, paper-based data capture requires transporting and retrieving the data captured on hard copies, followed by subsequent (double) data entry. EDC, therefore, results in a much faster availability of data for monitoring and analysis without time-delays and limited need for verification towards raw data.

While the setup of EDC studies may require more technical expertise and planning, the requirements for the use of EDC on the side of the end-user are relatively easy to fulfil. On the one hand, the use of EDC requires an end device with an ideally fast and reliable internet connection and an up-to-date internet browser to access the EDC system. Consequently, it is obligatory that an EDC tool is compatible with all commonly used web browsers. On the other hand, EDC users, i.e. all study participants entering data, need to have at least basic computer knowledge and must be willing to adapt to this "modern" way of data capture.

As the use of EDC has become more common in the animal health industry, regulatory authorities in Europe and the US have defined standards for the use of EDC systems in clinical studies. In general, the legal requirements regarding the accuracy, integrity, and correctness of data are the same for both paper-based and electronic records. The VICH GL9 Guideline on Good Clinical Practices (GCP) applies worldwide and is integrated in European legislation. According to the definition of the GCP guideline, raw data are any original records and first-hand observations made during a study. Consequently, when data are entered directly into an EDC system, the electronic record is considered the raw data. This means that per se usually no raw data in paper form exist if data are recorded contemporaneously into the EDC system; but certainly raw data in paper form might exist for some kind of data, like owner consents, lab sheets, owner diaries, records on backup forms and further contact or practice records. Similar to EU regulation, Title 21 CRF Part 11 published by the Food and Drug Administration (FDA) applies in the US. This regulation defines the criteria under which electronic records are accepted by the FDA to be equivalent to paper records in terms of trustworthiness and reliability. In addition, the FDA released several supplemental guidance documents that provide recommendations how these criteria shall be realised. According to these guidelines, only authorised individuals should be allowed to access the EDC system. This authorisation can be accomplished by providing each user an individual account with a unique username and password. The security of passwords may be enhanced by enabling the administrator to determine minimum length and composition of passwords. In addition, some EDC systems offer two-factor authentication to make the login process even more secure. The FDA further expects any EDC system to block user accounts after a defined number of failed login attempts and requires passwords to be changed on a regular basis. Study participants are usually provided with their login information via automatic emails sent by the system. Before users receive access at study start, they are required to have sufficient training in using the EDC system according to FDA. Such training should be carried out by qualified personnel and address not only the features of the system itself, but also data security aspects and the correct entry of data. All users should be trained on the study-specific requirements and the training should be documented. As soon as the active study phase is finished or at the completion of a study, access to the study database should be removed for all participants.

Regarding access rights, some EDC systems have implemented features to define different user roles. Those user roles will hold specific rights for processing data within the EDC system. Some user roles may allow entering data, while others will be read-only, i.e. only allowed to view but not

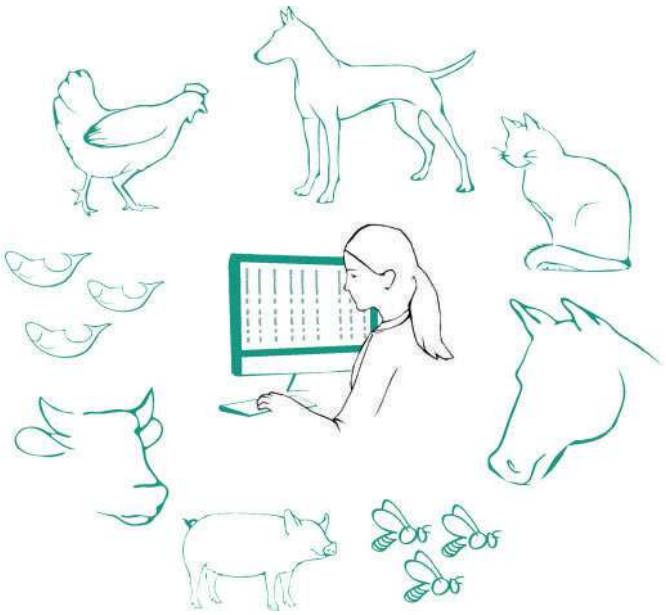

Figure 1. Ideally, EDC systems offer the flexibility to handle the different data generated in the broad variety of veterinary clinical studies, be it data from individual animals like dogs or cats, horses or dairy cows, or group data from animals as different as swine, poultry, or even aquaculture and bees.

to modify data, like monitors. Moreover, it can be assigned which visits and forms each role should have access to and whether the role needs the right to monitor or sign data. It is also possible to define which user roles need to be blinded vs. e.g. treatment information and which roles will stay unblinded, like for example dispensers administering treatments. We recommend specifying and defining the different roles in the data management plan and validating each user role specifically at the time when the EDC system is set up, as well as whether study participants are assigned to the correct user role. A list of all study participants and their roles has to be held in place as part of the study documentation.

Just like in paper-based studies, electronic raw data must be recorded on pre-established data capture forms to comply with the GCP standards (Figure 1). The pre-defined format and content of such data records not only promotes that data are collected according to the protocol requirements, but also that data are collected objectively and in the same way for all different individuals in each (treatment) group in the study. When implementing DCFs in EDC, one should consider programming the system to allow specific data entry fields to be completed automatically, like the patient ID or visit, to reduce the risk of entry errors. We further recommend including flags to mark data as transcribed and to add fields to record the date of and the reason for transcription (Figure 2). Calculated fields can be helpful to generate cross-references, e.g. to display data entered on one form also on another form, if useful.

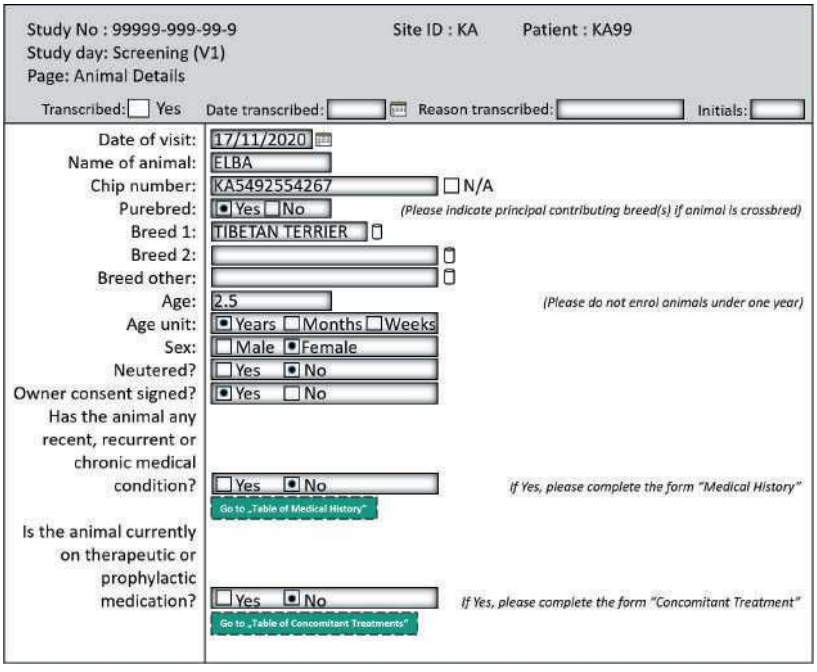

Figure 2. Example of an electronic data capture form (eDCF).

For all points in time when patients might possibly visit or be visited by the investigator, a section for data collection has to be made available, including all forms to be completed at that visit. This may include planned, optional, and unplanned visits. It further needs to be checked against the protocol whether the schedule is fixed or needs a dynamic sort order, e.g. when further study procedures will differ between arms or depend on therapy success. Determination of different arms or branches is possible in some EDC systems. Depending on the EDC system, study visits will be displayed to the user as grid or table or alternatively in a tree structure.

Two options exist to add data collected by the animal owner to the EDC database, like owner diaries or questionnaires. Such data can either be collected on paper or nowadays also via so-called electronic patient-reported outcomes (ePRO), a digital tool for data collection to be used by animal owners. Both options have their advantages and disadvantages. While paper-based owner diaries will need to be transcribed, verified and stored as raw data, ePROs require the owner to be willing and capable to enter data via a (mobile) device. Eventually, such owner diary forms can be assigned to the respective visit or be pooled all together in a dedicated section.

Laboratory data, such as blood or urine results, are usually part of the safety and efficacy observations and consequently the statistical analysis of a clinical study. Ideally, an EDC system allows importing such data into the database so that all data are available and stored in one place. Various aspects should be addressed, including how the original files will look, how often files should be imported and what data need to be imported; this should be done already during the planning phase of a study to be able to set up an adequate import routine that has been validated. If electronic lab files are intended to be imported, it can be helpful to ask the lab for a template file.

EDC can be particularly challenging in farm animal studies, where data may be collected in remote areas or in a farm barn with no internet connection. In such cases, offline systems may be useful where data upload to the central database will be done as soon as an internet connection is established. Non-patient-specific information, like e.g. environmental data, yield data or information about the farms, may be included in non-patient-specific study sections, a functionality available for some EDC systems.

Perhaps one of the most useful components of EDC systems, but on the same side most critical in terms of unblinding to treatment, is a randomisation tool. In the EDC systems, different types of randomisation options are available to randomly allocate patients to treatment groups. Most of the EDC systems provide built-in randomisation algorithms or offer to upload a pre-defined randomisation list. Randomisation via lists will allocate the patient to the treatment based on the randomisation criteria defined in that list. Direct randomisation in EDC systems may not always be the method of choice depending on the underlying randomisation algorithm. It is highly recommended to check if the allocation of patients to treatment groups may be unbalanced for small to medium patient numbers. Usually, a user of the defined user role triggers the randomisation by clicking a button or completing a key field, e.g. the question of whether the animal is suitable for the study is confirmed. To ensure randomisation is possible even in case the EDC system is not available, e.g. due to internet failure, we strongly recommend providing the study site(s) with an alternative to proceed, e.g. in case of an uploaded list provide a sealed printout of the randomisation list.

Before a system can be released to be ready to use, both the EDC system and the study-specific setup need to be validated and undergo stringent user acceptance testing. Standard operating procedures (SOPs) should be established to define the tests and measures for approval, as well as defining how and where the test results are going to be documented. Once the study has gone live, thus is active, every update of the EDC system and any modification to the

study-specific setup requires re-validation and appropriate documentation. Documentation of re-validation should include a detailed description of the intended changes, date and results of the validation, potential risks and finally a date of release for use, as far as applicable.

As for data collected on paper, an EDC system should allow keeping track of all changes made to the raw data. Any data modifications, additions, or deletions must be traceable and stored in the EDC system in the form of an audit trail. This audit trail enables the reconstruction of the original data at any time. Elements that should be included in the audit trail are information on the user modifying the data, the date and time when the modification took place, the new and the old value, as well as the reason for the modification. The point in time when data is saved to the audit trail depends on the EDC system. Data can either be saved after a form has been entered completely and left, after a "safe" button has been clicked or after an individual data entry field has been completed and left to enter the next one. The disadvantage of "safe" buttons is that users might forget to click and in consequence data inadvertently get lost and need to be re-entered. On the contrary, if data are immediately stored when a field is left, there might be many entry errors. Therefore, some EDC systems offer the possibility to define a time lapse during which the user will be able to modify data without adding any records to the audit trail. However, we do not recommend using this option since FDA compliance is not guaranteed.

With the use of EDC systems, we consider that a higher level of data quality is reached. Most systems allow configuring electronic checks for completeness, consistency and plausibility of data. When incorrect data are entered or missing overall, the system will highlight entry errors. These electronic checks can be conducted subsequently after data entry of a form by performing so-called edit checks or pre-tests, or on a regular basis in the form of messages or queries. More detailed and complex checks not covered by simple electronic checks may be accomplished via additional queries in the respective database language or by monitors or medical supervisors who review the data and write manual queries. Even though electronic checks are a great way to improve data quality, they should not prevent or restrict any data from being entered.

In most EDC systems, after all queries have been resolved, the investigator has to sign the eDCF electronically. According to the FDA guideline, such electronic signatures have to contain the name of the signer, the date and time of, and the reason for, signature. Additionally, user authorisation for the electronic signature must be unique to the user and must not be used by anyone else. In order to sign a form electronically within the EDC system, users are asked to enter their personal login and password.

Any data recorded in other languages than in English may require to be translated. Depending on the EDC system used, translations may be imported or entered manually into the eDCFs. Common or recurrent entries can be translated in a standardised way via coding lists. Coding improves consistency of data, since particular terms will always be coded with the same wording. In case official coding dictionaries are used (e.g. VeDDRA, ATCvet), the version of the coding dictionary should be documented. Besides contributing to the standardisation of terms, coding can also serve a regulatory purpose (e.g. in pharmacovigilance), or be used to mask data such as treatment groups in case of blinding or personal data of the animal owners for data privacy protection reasons.

Once all activities are completed and data are clean, a data pack will be prepared in consultation with those who are going to receive it. Before data are extracted, the database has to be closed by a database lock. This prevents any further changes being made to the database. A data pack must only be prepared from the final database after the database has been locked. The data pack will have to be exported in a format appropriate for further processing by the statistician. Usually, EDC systems will offer to export data to SAS, XML or other common formats.

Eventually, the database will be archived. According to GCP, all study documentation including the raw data must be retained for at least the period of time dictated by the relevant regulatory authorities. Since almost all study data only exist in electronic form in EDC studies, it is vital to implement sufficient backup and recovery procedures that protect against data loss – for both during and after conduct of the study.

In conclusion, EDC systems are by now probably the most convenient method to gather patient data in clinical trials. Real-time access to data, increased efficiency and cleaner data are one of the main reasons why the majority of clinical studies nowadays use EDC instead of paper-based methods. EDC systems offer a variety of solutions and features for different kinds of data management processes, including but not limited to data entry, plausibility checks, monitoring, coding and import/export of data. These processes need to go through careful planning and specification before setting them up in EDC based on the study-specific requirements. In fact, we highly recommend summarising all relevant data management tasks in a data management plan beforehand. This will not only allow for more reliability and transparency of the defined tasks, but will also help to exploit EDC systems to their full potential.

Katharina Hiemer

Katharina is a biologist with an academic background in marine biology. Before diving into data management, she worked in scientific research at the Alfred Wegener Institute for Polar and Marine Research. Her research work involved analysing long-term underwater recordings to investigate the temporal distribution and vocal behaviour of marine mammals in polar oceans. In 2020, she joined KLIFOVET AG as Clinical Data Manager.

Email: katharina.hiemer@klifovet.com

Brigitte Steiner

Brigitte is a veterinarian with more than 16 years' experience in the animal health industry. After several years in veterinary practice, she joined KLIFOVET AG as a project manager for veterinary clinical studies in 2004. Six years later she moved into data management and focused on data management procedures, development and validation of EDC systems specifically developed for animal health. Since 2017, Brigitte has been heading the Data Management and IT department at KLIFOVET AG.

Email: brigitte.steiner@klifovet.com